

Asynchronous Adaptation and Learning over Networks — Part II: Performance Analysis

Xiaochuan Zhao, *Student Member, IEEE*, and Ali H. Sayed, *Fellow, IEEE*

Abstract

In Part I [2], we introduced a fairly general model for asynchronous events over adaptive networks including random topologies, random link failures, random data arrival times, and agents turning on and off randomly. We performed a stability analysis and established the notable fact that the network is still able to converge in the mean-square-error sense to the desired solution. Once stable behavior is guaranteed, it becomes important to evaluate how fast the iterates converge and how close they get to the optimal solution. This is a demanding task due to the various asynchronous events and due to the fact that agents influence each other. In this Part II, we carry out a detailed analysis of the mean-square-error performance of asynchronous strategies for solving distributed optimization and adaptation problems over networks. We derive analytical expressions for the mean-square convergence rate and the steady-state mean-square-deviation. The expressions reveal how the various parameters of the asynchronous behavior influence network performance. In the process, we establish the interesting conclusion that even under the influence of asynchronous events, all agents in the adaptive network can still reach an $O(\nu^{1+\gamma'_o})$ near-agreement with some $\gamma'_o > 0$ while approaching the desired solution within $O(\nu)$ accuracy, where ν is proportional to the small step-size parameter for adaptation.

Index Terms

Distributed learning, distributed optimization, diffusion adaptation, asynchronous behavior, adaptive networks, dynamic topology, link failures.

The authors are with Department of Electrical Engineering, University of California, Los Angeles, CA 90095 Email: {xzhao, sayed}@ee.ucla.edu.

This work was supported by NSF grants CCF-1011918 and ECCS-1407712. A short and limited early version of this work appeared in the conference proceeding [1]. The first part of this work is presented in [2].

I. INTRODUCTION

In Part I [2], we introduced a fairly general model for asynchronous distributed adaptation and learning over networks. The model allows the step-size for any agent to be randomly chosen within a range of values including zero, meaning an off-status; it also allows the communication links between any two agents to be randomly turned on and off; it further allows the topology to vary and the combination weights on the links between agents to vary randomly. The model also captures the situation in which agents randomly select a subset of their neighbors to share information with, as happens in gossip implementations. Based on this asynchronous model, we carried out a detailed stability analysis in Part I [2] and arrived at Theorem 1 in that Part, which provides an explicit condition on the first and second-order moments of the distribution of the step-sizes to ensure mean-square stability of the adaptive network. When the condition holds, it was further shown that the asymptotic mean-square-deviation (MSD) for every agent in the network remains bounded. Interestingly, it was shown that these conclusions hold irrespective of the randomness in the network topology.

In this Part II, we conduct a detailed mean-square-error (MSE) analysis in order to characterize the learning behavior of the asynchronous network in terms of the network parameters. In particular, we answer the following three questions:

- How is the convergence rate of the algorithm affected by the occurrence of asynchronous events in comparison to a synchronous network?
- Are agents still able to reach some sort of agreement in steady-state despite the various sources of randomness influencing their interactions?
- How close do the iterates of the various agents get to each other and to the desired optimal solution?

One of the main conclusions that will follow from the analysis is that, under certain reasonable conditions, the asynchronous network will continue to be able to deliver performance that is comparable to the synchronous case where no failures occur. In particular, we will be able to establish that for a sufficiently small step-size parameter ν it holds that

$$\lim_{i \rightarrow \infty} \mathbb{E} \|w^o - \mathbf{w}_{k,i}\|^2 = O(\nu) \quad (1)$$

$$\lim_{i \rightarrow \infty} \mathbb{E} \|\mathbf{w}_{k,i} - \mathbf{w}_{\ell,i}\|^2 = O(\nu^{1+\gamma'_o}) \quad (2)$$

for some $\gamma'_o > 0$ that is given by (92) further ahead. These results imply that all agents reach a level of $O(\nu^{1+\gamma'_o})$ agreement with each other and get $O(\nu)$ close to the desired optimal solution, w^o , in steady-state. This interesting behavior is illustrated in Fig. 1, where it is shown that, despite being subjected

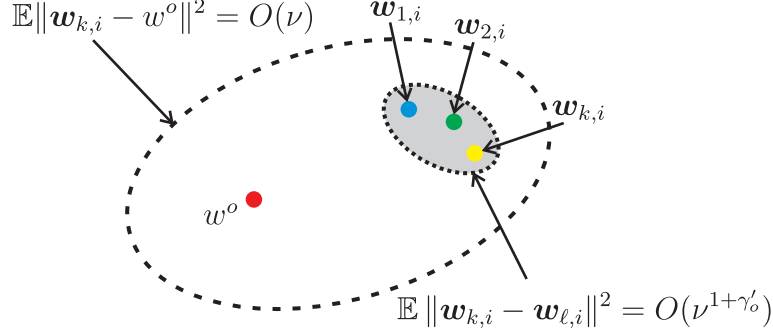


Fig. 1. Illustration for Eqs. (1) and (2): the agents do not only get $O(\nu)$ close to the target w^o but they also cluster next to each other within $O(\nu^{1+\gamma'_o})$ for some $\gamma'_o > 0$.

to various sources of randomness, the agents in an asynchronous network are still able to approach the desired solution and they are also able to coalesce close to each other. For the remainder of this part, we continue to use the same symbols, notation, and modeling assumptions from Part I [2]. Moreover, we focus on presenting the main results and their interpretation in the body of the paper, while delaying the detailed proofs and arguments to the appendices.

Notation: We use lowercase letters to denote vectors, uppercase letters for matrices, plain letters for deterministic variables, and boldface letters for random variables. We also use $(\cdot)^\top$ to denote transposition, $(\cdot)^*$ to denote conjugate transposition, $(\cdot)^{-1}$ for matrix inversion, $\text{Tr}(\cdot)$ for the trace of a matrix, $\lambda(\cdot)$ for the eigenvalues of a matrix, $\|\cdot\|$ for the 2-norm of a matrix or the Euclidean norm of a vector, and $\rho(\cdot)$ for the spectral radius of a matrix. Besides, we use \otimes to denote the Kronecker product and \otimes_b to denote the block Kronecker product.

II. NETWORK ERROR DYNAMICS

In Part I [2], we examined the mean-square stability of the following asynchronous diffusion strategy:

$$\psi_{k,i} = \mathbf{w}_{k,i-1} - \mu_k(i) \widehat{\nabla_{w^*} J_k}(\mathbf{w}_{k,i-1}) \quad (3a)$$

$$\mathbf{w}_{k,i} = \sum_{\ell \in \mathcal{N}_{k,i}} \mathbf{a}_{\ell k}(i) \psi_{\ell,i} \quad (3b)$$

for the distributed solution of a global inference task in the following form:

$$\underset{w}{\text{minimize}} \quad J^{\text{glob}}(w) \triangleq \sum_{k=1}^N J_k(w) \quad (4)$$

The global optimum of (4) is denoted by w^o . We continue to assume the same asynchronous model from Part I [2]. Specifically, the random step-sizes $\{\mu_k(i)\}$ and random combination coefficients $\{\mathbf{a}_{\ell k}(i)\}$ in (3a)–(3b) satisfy the conditions described in Section III-B of Part I [2]; and the cost functions $\{J_k(w)\}$

satisfy Assumptions 1 and 2 in Section II of Part I [2]. However, in order to study the mean-square-error performance in *steady-state*, it is necessary to strengthen the assumption on the stochastic gradient vectors $\{\widehat{\nabla_{w^*} J_k}(\mathbf{w}_{k,i-1})\}$. We replace the gradient noise model described in Assumption 3 in Section III-A of Part I [2] by the following one.

Assumption 1 (Gradient noise model):

- 1) The gradient noise $\mathbf{v}_{k,i}(\mathbf{w}_{k,i-1})$, conditioned on \mathbb{F}_{i-1} , is assumed to be independent of any other random sources including topology, links, combination coefficients, and step-sizes. The conditional moments of $\mathbf{v}_{k,i}(\mathbf{w}_{k,i-1})$ satisfy:

$$\mathbb{E}[\mathbf{v}_{k,i}(\mathbf{w}_{k,i-1})|\mathbb{F}_{i-1}] = 0 \quad (5)$$

$$\mathbb{E}[\|\mathbf{v}_{k,i}(\mathbf{w}_{k,i-1})\|^4|\mathbb{F}_{i-1}] \leq \alpha^2\|w^o - \mathbf{w}_{k,i-1}\|^4 + \sigma_v^4 \quad (6)$$

for some $\alpha \geq 0$ and $\sigma_v^2 \geq 0$.

- 2) The individual gradient noises $\{\mathbf{v}_{k,i}(\mathbf{w}_{k,i-1})\}$ are uncorrelated and circular across all agents such that

$$\mathcal{R}_i(\mathbf{w}_{i-1}) = \text{diag}\{R_{1,i}(\mathbf{w}_{1,i-1}), \dots, R_{N,i}(\mathbf{w}_{N,i-1})\} \quad (7)$$

where $\mathcal{R}_i(\mathbf{w}_{i-1})$ and $\{R_{k,i}(\mathbf{w}_{k,i-1})\}$ are from (23) and (19) both in Part I [2].

- 3) The conditional covariance of $\mathbf{v}_i(\mathbf{w}_{i-1})$ satisfies the Lipschitz condition

$$\|\mathcal{R}_i(\mathbf{1}_N \otimes w^o) - \mathcal{R}_i(\mathbf{w}_{i-1})\| \leq \kappa_v \|\mathbf{1}_N \otimes w^o - \mathbf{w}_{i-1}\|^{\gamma_v} \quad (8)$$

for some constants $\kappa_v \geq 0$ and $0 < \gamma_v \leq 4$.

- 4) The covariance of $\mathbf{v}_i(\mathbf{1}_N \otimes w^o)$ converges to a constant matrix:

$$\mathcal{R} \triangleq \lim_{i \rightarrow \infty} \mathcal{R}_i(\mathbf{1}_N \otimes w^o) \triangleq \text{diag}\{R_1, \dots, R_N\} \quad (9)$$

where

$$R_k \triangleq \lim_{i \rightarrow \infty} R_{k,i}(w^o) \quad (10)$$

■

From Assumption 1, the conditional moments of $\mathbf{v}_{k,i}(\mathbf{w}_{k,i-1})$ satisfy

$$\mathbb{E}[\mathbf{v}_{k,i}(\mathbf{w}_{k,i-1})|\mathbb{F}_{i-1}] = 0 \quad (11)$$

$$\mathbb{E}[\|\mathbf{v}_{k,i}(\mathbf{w}_{k,i-1})\|^4|\mathbb{F}_{i-1}] \leq \alpha^2\|\underline{w}^o - \underline{\mathbf{w}}_{k,i-1}\|^4 + 4\sigma_v^4 \quad (12)$$

where a factor of 4 appeared due to the transform $\mathbb{T}(\cdot)$ from (4) of Part I [2].

A. Long Term Error Dynamics

We showed in (87) from Part I [2] that the error recursion for the asynchronous network (3a)–(3b) evolves according to the following dynamics:

$$\tilde{\mathbf{w}}_i = \mathcal{A}_i^\top (I_{2MN} - \mathcal{M}_i \mathcal{H}_{i-1}) \tilde{\mathbf{w}}_{i-1} + \mathcal{A}_i^\top \mathcal{M}_i \mathbf{v}_i(\mathbf{w}_{i-1}) \quad (13)$$

where $\mathcal{H}_{i-1} = \text{diag}\{\mathbf{H}_{1,i-1}, \mathbf{H}_{2,i-1}, \dots, \mathbf{H}_{N,i-1}\}$ and

$$\mathbf{H}_{k,i-1} \triangleq \int_0^1 \nabla_{\underline{\mathbf{w}}}^2 J_k(\underline{\mathbf{w}}^o - t \tilde{\mathbf{w}}_{k,i-1}) dt \quad (14)$$

The dependency of \mathcal{H}_{i-1} on the previous iterate \mathbf{w}_{i-1} complicates the mean-square analysis. Recall though from Lemma 1 in Part I [2] that the Hessian matrices of the costs $\{J_k(\underline{\mathbf{w}})\}$ are globally Lipschitz around $\underline{\mathbf{w}}^o$. Let

$$\mathbf{H}_k \triangleq \nabla_{\underline{\mathbf{w}}}^2 J_k(\underline{\mathbf{w}}^o), \quad \mathcal{H} \triangleq \text{diag}\{\mathbf{H}_1, \dots, \mathbf{H}_N\} \quad (15)$$

Recursion (13) can then be rewritten as

$$\tilde{\mathbf{w}}_i = \mathcal{A}_i^\top (I_{2MN} - \mathcal{M}_i \mathcal{H}) \tilde{\mathbf{w}}_{i-1} + \mathcal{A}_i^\top \mathcal{M}_i \mathbf{v}_i(\mathbf{w}_{i-1}) + \mathcal{A}_i^\top \mathbf{d}_i \quad (16)$$

where the perturbation factor \mathbf{d}_i is given by

$$\mathbf{d}_i \triangleq \mathcal{M}_i (\mathcal{H} - \mathcal{H}_{i-1}) \tilde{\mathbf{w}}_{i-1} \triangleq \text{col}\{\mathbf{d}_{1,i}, \dots, \mathbf{d}_{N,i}\} \quad (17)$$

$$\mathbf{d}_{k,i} \triangleq \boldsymbol{\mu}_k(i) (\mathbf{H}_k - \mathbf{H}_{k,i-1}) \tilde{\mathbf{w}}_{k,i-1} \quad (18)$$

Let $\bar{\mu}_k^{(n)} \triangleq \mathbb{E}[\boldsymbol{\mu}_k(i)]^n$ denote the n -th moment of the random step-size parameter $\boldsymbol{\mu}_k(i)$; we also use $\bar{\mu}_k \equiv \bar{\mu}_k^{(1)}$ from (34) of Part I [2] for the mean and $c_{\mu,k,\ell} = \mathbb{E}[(\boldsymbol{\mu}_k(i) - \bar{\mu}_k)(\boldsymbol{\mu}_\ell(i) - \bar{\mu}_\ell)]$ from (37) of Part I [2] for the cross-covariance.

Lemma 1 (Size of perturbation): If condition (103) in Part I [2], namely,

$$\frac{\sqrt{\bar{\mu}_k^{(4)}}}{\bar{\mu}_k^{(1)}} < \frac{\lambda_{k,\min}}{3\lambda_{k,\max}^2 + 4\alpha} \quad (19)$$

holds for all k , then

$$\limsup_{i \rightarrow \infty} \mathbb{E} \|\mathcal{A}_i^\top \mathbf{d}_i\|^2 \leq O(\nu^4) \quad (20)$$

where

$$\nu \triangleq \max_k \frac{\sqrt{\bar{\mu}_k^{(4)}}}{\bar{\mu}_k^{(1)}} \quad (21)$$

Proof: See Appendix A. ■

Assumption 2 (Small step-sizes): The parameter ν from (21) is sufficiently small such that

$$\nu < \min_k \frac{\lambda_{k,\min}}{3\lambda_{k,\max}^2 + 4\alpha} < 1 \quad (22)$$

■

Under Assumption 2, condition (19) holds. It was shown in (200) and (201) from Part I [2] that condition (19) in this Part implies condition (93) from Part I [2], i.e.,

$$\frac{\bar{\mu}_k^{(2)}}{\bar{\mu}_k^{(1)}} < \frac{\lambda_{k,\min}}{\alpha + \lambda_{k,\max}^2} \quad (23)$$

for all k .

Since we are interested in examining the asymptotic performance of the asynchronous network, result (20) indicates that the network error recursion (13) can be expressed for large enough i by using the following *long-term* model:

$$\tilde{\mathbf{w}}'_i = \mathcal{A}_i^\top (I_{2MN} - \mathcal{M}_i \mathcal{H}) \tilde{\mathbf{w}}'_{i-1} + \mathcal{A}_i^\top \mathcal{M}_i \mathbf{v}_i(\mathbf{w}_{i-1}) \quad (24)$$

where we ignore the $O(\nu^2)$ term $\mathcal{A}_i^\top \mathbf{d}_i$ according to (20), and we use \mathbf{w}'_{i-1} to denote the estimate obtained from this long-term model. It is worth noting that the gradient noise $\mathbf{v}_i(\mathbf{w}_{i-1})$ in (24) is an extraneous noise that is imported from the original model (13); it only depends on the original estimate \mathbf{w}_{i-1} but not on \mathbf{w}'_{i-1} . We will now use recursion (24) to determine expressions (rather than bounds) for the steady-state individual MSD and for the average network MSD. One advantage of model (24) is that the random matrix \mathcal{H}_{i-1} from (13) has been replaced by the constant matrix \mathcal{H} . More formally, under Assumption 1 on the fourth-order moment of the gradient noise, and by extending the arguments of Appendices D and E from Part I [2] and the arguments of [3], we will establish later in (76) that the MSD expression resulting from (24) is within $O(\nu^{3/2})$ of the MSD expression for the original recursion (13); this conclusion will rely on the following useful result.

Theorem 1 (Bounded mean-square gap): Under Assumptions 1 and 2, the mean-square gap from the original error recursion (13) to the long-term model (24) is then asymptotically bounded by

$$\limsup_{i \rightarrow \infty} \left[\max_k \mathbb{E} \|\tilde{\mathbf{w}}_{k,i} - \tilde{\mathbf{w}}'_{k,i}\|^2 \right] \leq O(\nu^2) \quad (25)$$

for any k .

Proof: See Appendix B. ■

To proceed with the mean-square-error performance analysis, we introduce the following auxiliary variables:

$$\mathbf{D}_{k,i} \triangleq I_{2M} - \boldsymbol{\mu}_k(i) H_k \quad (26)$$

$$\mathcal{D}_i \triangleq I_{2MN} - \mathcal{M}_i \mathcal{H} = \text{diag}\{\mathcal{D}_{k,i}\} \quad (27)$$

$$\mathcal{B}_i \triangleq \mathcal{A}_i^\top \mathcal{D}_i \quad (28)$$

$$\mathbf{s}_i \triangleq \mathcal{A}_i^\top \mathcal{M}_i \mathbf{v}_i(\mathbf{w}_{i-1}) \quad (29)$$

Based on the gradient noise model in Assumption 1 and the asynchronous network model described in Section III-B of Part I [2], it is easy to verify that the (conditional) means of $\{\mathcal{A}_i, \mathcal{M}_i, \mathcal{D}_{k,i}, \mathcal{D}_i, \mathcal{B}_i, \mathbf{s}_i\}$ are given by:

$$\bar{\mathcal{A}} \triangleq \mathbb{E}(\mathcal{A}_i) = \bar{A} \otimes I_{2M} \quad (30)$$

$$\bar{\mathcal{M}} \triangleq \mathbb{E}(\mathcal{M}_i) = \bar{M} \otimes I_{2M} \quad (31)$$

$$\bar{\mathcal{D}}_k \triangleq \mathbb{E}(\mathcal{D}_{k,i}) = I_{2M} - \bar{\mu}_k H_k \quad (32)$$

$$\bar{\mathcal{D}} \triangleq \mathbb{E}(\mathcal{D}_i) = I_{2MN} - \bar{\mathcal{M}} \mathcal{H} = \text{diag}\{\bar{\mathcal{D}}_k\} \quad (33)$$

$$\bar{\mathcal{B}} \triangleq \mathbb{E}(\mathcal{B}_i) = \bar{\mathcal{A}}^\top \bar{\mathcal{D}} \quad (34)$$

$$\bar{\mathbf{s}} \triangleq \mathbb{E}(\mathbf{s}_i | \mathbb{F}_{i-1}) = 0 \quad (35)$$

It can be verified that the block-Kronecker-covariance matrices of several random quantities are given by:

$$\mathcal{C}_A \triangleq \mathbb{E}[(\mathcal{A}_i - \bar{\mathcal{A}}) \otimes_b (\mathcal{A}_i - \bar{\mathcal{A}})] = C_A \otimes I_{4M^2} \quad (36)$$

$$\mathcal{C}_M \triangleq \mathbb{E}[(\mathcal{M}_i - \bar{\mathcal{M}}) \otimes_b (\mathcal{M}_i - \bar{\mathcal{M}})] = C_M \otimes I_{4M^2} \quad (37)$$

$$\mathcal{C}_D \triangleq \mathbb{E}[(\mathcal{D}_i^* - \bar{\mathcal{D}}^*)^\top \otimes_b (\mathcal{D}_i - \bar{\mathcal{D}})] = C_M (\mathcal{H}^\top \otimes_b \mathcal{H}) \quad (38)$$

$$\mathcal{C}_B \triangleq \mathbb{E}[(\mathcal{B}_i^* - \bar{\mathcal{B}}^*)^\top \otimes_b (\mathcal{B}_i - \bar{\mathcal{B}})] = (\bar{\mathcal{A}}^\top \otimes_b \bar{\mathcal{A}}^\top) \mathcal{C}_D + \mathcal{C}_A^\top (\bar{\mathcal{D}}^\top \otimes_b \bar{\mathcal{D}} + \mathcal{C}_D) \quad (39)$$

where the symbol \otimes_b denotes the block-Kronecker operation of block size $2M \times 2M$ (see Appendix C).

Moreover, it can be verified by using property (133) from Appendix C that

$$\mathbb{E}[(\mathcal{X}^* - \bar{\mathcal{X}}^*)^\top \otimes_b (\mathcal{X} - \bar{\mathcal{X}})] = \mathbb{E}[(\mathcal{X}^*)^\top \otimes_b \mathcal{X}] - (\bar{\mathcal{X}}^*)^\top \otimes_b \bar{\mathcal{X}} \quad (40)$$

for any random block matrix \mathcal{X} with appropriate block size and with mean $\bar{\mathcal{X}} \triangleq \mathbb{E}\mathcal{X}$. The $\{C_A, C_M\}$ that appear in (36)–(39) relate to the second-order moments of $\{\mathbf{a}_{\ell k}(i)\}$ and $\{\boldsymbol{\mu}_k(i)\}$. Using (28) and (29), the long-term model (24) can be rewritten as

$$\tilde{\mathbf{w}}'_i = \mathcal{B}_i \cdot \tilde{\mathbf{w}}'_{i-1} + \mathbf{s}_i \quad (41)$$

B. Mean Error Recursion

Taking the expectation of both sides of (41), we end up with the mean error recursion for large i :

$$\mathbb{E} \tilde{\mathbf{w}}'_i = \bar{\mathcal{B}} \cdot \mathbb{E} \tilde{\mathbf{w}}'_{i-1} \quad (42)$$

The stability of recursion (42) requires the stability of $\bar{\mathcal{B}}$. A condition on the step-sizes to ensure the stability of $\bar{\mathcal{B}}$ can be derived as follows. Using the fact that $\bar{\mathcal{A}}$ is block left-stochastic and $\bar{\mathcal{D}}$ is block diagonal and Hermitian, and following the same argument in [4, App. A] [5], we obtain

$$\rho(\bar{\mathcal{B}}) \leq \rho(\bar{\mathcal{D}}) \quad (43)$$

where $\rho(\cdot)$ denotes the spectral radius of its matrix argument. It follows from (33) and (43) that asymptotic mean stability is guaranteed if the mean step-size $\bar{\mu}_k$ satisfies

$$\bar{\mu}_k \equiv \bar{\mu}_k^{(1)} < \frac{2}{\rho(H_k)} \quad (44)$$

for all k . Since H_k is a positive semi-definite matrix, its spectral radius coincides with its largest eigenvalue. Using (8) from Part I [2], we have $\rho(H_k) \leq \lambda_{k,\max}$. If condition (19) holds, then from (196) of Part I [2], we have

$$\bar{\mu}_k^{(1)} \leq \frac{\sqrt{\bar{\mu}_k^{(4)}}}{\bar{\mu}_k^{(1)}} < \frac{\lambda_{k,\min}}{3\lambda_{k,\max}^2 + 4\alpha} \leq \frac{\lambda_{k,\min}}{3\lambda_{k,\max}^2} \leq \frac{2}{\rho(H_k)} \quad (45)$$

since $\alpha > 0$. Therefore, condition (44) holds if condition (19) does so. With Assumption 2, we have

$$\lim_{i \rightarrow \infty} \mathbb{E} \tilde{\mathbf{w}}'_{k,i} = 0 \quad (46)$$

for all k . From (46), we conclude that the long-term model (24) or, equivalently, (41), is the asymptotically centered version of the original error recursion (13).

C. Error Covariance Recursion

We proceed to examine the evolution of the covariance matrix of the network error vector $\tilde{\mathbf{w}}'_i$ in the long-term model (41). Let

$$r_i(\mathbf{w}_{i-1}) \triangleq \text{bvec}(\mathcal{R}_i(\mathbf{w}_{i-1})) = \mathbb{E}[(\underline{\mathbf{v}}_i^*(\mathbf{w}_{i-1}))^\top \otimes_b \underline{\mathbf{v}}_i(\mathbf{w}_{i-1}) | \mathbb{F}_{i-1}] \quad (47)$$

$$\mathbf{y}_i \triangleq (\bar{\mathcal{A}}^\top \otimes_b \bar{\mathcal{A}}^\top + \mathcal{C}_A^\top)(\bar{\mathcal{M}} \otimes_b \bar{\mathcal{M}} + \mathcal{C}_M) \mathbb{E}[r_i(\mathbf{w}_{i-1})] \quad (48)$$

$$\mathbf{z}_i \triangleq \text{bvec}(\mathbb{E}(\tilde{\mathbf{w}}'_i \tilde{\mathbf{w}}'^{*}_i)) = \mathbb{E}[(\tilde{\mathbf{w}}'^{*}_i)^\top \otimes_b \tilde{\mathbf{w}}'_i] \quad (49)$$

$$\mathcal{G} \triangleq \mathbb{E}[(\mathcal{D}_i^*)^\top \otimes_b \mathcal{D}_i] = \bar{\mathcal{D}}^\top \otimes_b \bar{\mathcal{D}} + \mathcal{C}_D \quad (50)$$

$$\mathcal{F} \triangleq \mathbb{E}[(\mathcal{B}_i^*)^\top \otimes_b \mathcal{B}_i]^* = \bar{\mathcal{B}}^\top \otimes_b \bar{\mathcal{B}}^* + \mathcal{C}_B^* = \mathcal{G}(\bar{\mathcal{A}} \otimes_b \bar{\mathcal{A}} + \mathcal{C}_A) \quad (51)$$

where the notations $\text{bvec}(\cdot)$ and \otimes_b denote block vectorization and block Kronecker products, respectively, both of size $2M \times 2M$ (see Appendix C). We note that the second equalities in (47) and (49) are due to property (130) and the second equalities in (50) and (51) are by using (33), (34), and (38)–(40). Using (47)–(51), we obtain the following recursion for the block-vectorized covariance matrix of the network error vector $\tilde{\mathbf{w}}'_i$.

Theorem 2 (Network error covariance recursion): The vector z_i evolves according to the following recursion:

$$z_i = \mathcal{F}^* z_{i-1} + y_i \quad (52)$$

Recursion (52) converges if condition (19) holds, and its convergence rate is determined by $\rho(\mathcal{F})$.

Proof: See Appendix D. ■

The vector z_i can be used to compute useful error metrics. For example, we can examine any weighted MSE measure for $\tilde{\mathbf{w}}'_i$ by evaluating quantities of the form:

$$\mathbb{E} \|\tilde{\mathbf{w}}'_i\|_\Sigma^2 = \mathbb{E} [\text{Tr}(\tilde{\mathbf{w}}'_i \tilde{\mathbf{w}}'^{*}_i \Sigma)] = z_i^* \cdot \text{bvec}(\Sigma) \quad (53)$$

where Σ is an arbitrary positive semi-definite weight matrix. To guarantee the convergence of $\mathbb{E} \|\tilde{\mathbf{w}}'_i\|_\Sigma^2$ for any weighting matrix Σ , it is sufficient and necessary to guarantee the convergence of z_i . It follows from Theorem 2 that under Assumption 2, the spectral radius of the matrix \mathcal{F} in (52) determines the mean-square stability and convergence rate of the asynchronous diffusion strategy (3a)–(3b).

Before proceeding we comment on the reason why we choose to use the block vectorization operation $\text{bvec}(\cdot)$ in (49) instead of the traditional vectorization operation $\text{vec}(\cdot)$. This is because $\text{bvec}(\cdot)$ allows us to track each block of its matrix argument *after* vectorization. By the definition in (125) and the illustration in Fig. 2, operation $\text{bvec}(\cdot)$ preserves the locality of every block in the original matrix argument whereas operation $\text{vec}(\cdot)$ blends different blocks together. Therefore, whenever we need to vectorize a network matrix whose blocks relate to individual agents, it is more natural to use the block vectorization operation $\text{bvec}(\cdot)$; on the other hand, whenever we need to vectorize a matrix that only relates to a single agent, we can use the conventional vectorization operation $\text{vec}(\cdot)$. A useful property of the conventional vectorization operation $\text{vec}(\cdot)$ is

$$\text{vec}(ABC) = (C^\top \otimes A) \cdot \text{vec}(B) \quad (54)$$

for matrices $\{A, B, C\}$ of compatible sizes. A similar property holds for the $\text{bvec}(\cdot)$ operation:

$$\text{bvec}(ABC) = (C^\top \otimes_b A) \cdot \text{bvec}(B) \quad (55)$$

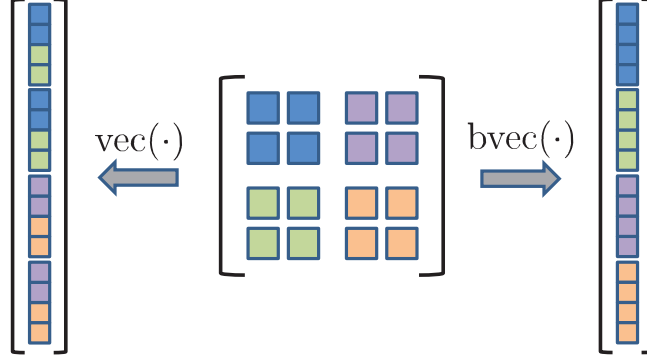


Fig. 2. Comparing two vectorization operations: $\text{vec}(\cdot)$ versus $\text{bvec}(\cdot)$. The operation $\text{vec}(\cdot)$ destroys the locality of the blocks in the original matrix argument while the operation $\text{bvec}(\cdot)$ preserves it.

for block matrices $\{A, B, C\}$ with appropriate block sizes. In Fig. 3, we compare the structures of $A \otimes B$ and $A \otimes_b B$, where $\{A, B\}$ are a pair of block matrices. The observation is that the operation \otimes destroys the locality of the blocks from matrix B , whereas the operation \otimes_b preserves the locality of the blocks from both matrices A and B .

Using properties of the block operations $\text{bvec}(\cdot)$ and \otimes_b , we can derive from Theorem 2 a useful relation between the blocks of the network error covariance matrix, $\mathbb{E}\tilde{\mathbf{w}}'_i \tilde{\mathbf{w}}^{*}_{i'}$, and the blocks of the vector z_i . Let us partition the $4M^2N^2$ -dimensional vector z_i as

$$z_i = \text{col}\{z_i^{(1)}, \dots, z_i^{(N)}\}, \quad z_i^{(\ell)} \triangleq \text{col}\{z_i^{(\ell,1)}, \dots, z_i^{(\ell,N)}\} \quad (56)$$

where $z_i^{(\ell)}$ is the ℓ -th sub-vector of z_i with dimension $4M^2N$ and $z_i^{(\ell,k)}$ is the k -th block of $z_i^{(\ell)}$ with dimension $4M^2$. From (49) and (125), we find that these vectors have the following useful interpretations for $k, \ell = 1, 2, \dots, N$:

$$z_i = \mathbb{E}[\text{bvec}(\tilde{\mathbf{w}}'_i \tilde{\mathbf{w}}^{*}_{i'})] = \text{col}\{\mathbb{E}[(\tilde{\mathbf{w}}^{*}_{\ell,i})^\top \otimes \tilde{\mathbf{w}}'_{k,i}]\}_{\ell,k=1}^N \quad (57)$$

$$z_i^{(\ell,k)} \triangleq \text{vec}(\mathbb{E}[\tilde{\mathbf{w}}'_{k,i} \tilde{\mathbf{w}}^{*}_{\ell,i}]) = \mathbb{E}[(\tilde{\mathbf{w}}^{*}_{\ell,i})^\top \otimes \tilde{\mathbf{w}}'_{k,i}] \quad (58)$$

where $\mathbb{E}\tilde{\mathbf{w}}'_{k,i} \tilde{\mathbf{w}}^{*}_{\ell,i}$ is the (k, ℓ) -th block of $\mathbb{E}\tilde{\mathbf{w}}'_i \tilde{\mathbf{w}}^{*}_{i'}$ with size $2M \times 2M$. The block entries of the vector z_i in (58) do not only allow us to recover the covariance matrices of any individual error vectors, $\mathbb{E}\tilde{\mathbf{w}}'_{k,i} \tilde{\mathbf{w}}^{*}_{k,i}$, but they also allow us to recover the *cross-covariance* matrices, $\mathbb{E}\tilde{\mathbf{w}}'_{k,i} \tilde{\mathbf{w}}^{*}_{\ell,i}$, for any pair of agents $\{k, \ell\}$. Therefore, by studying the evolution of the entire covariance vector in (52), we are able to extract some detailed information about the dynamics of the asynchronous diffusion network, as we shall show in Theorem 4 and Corollary 3 in Section IV.

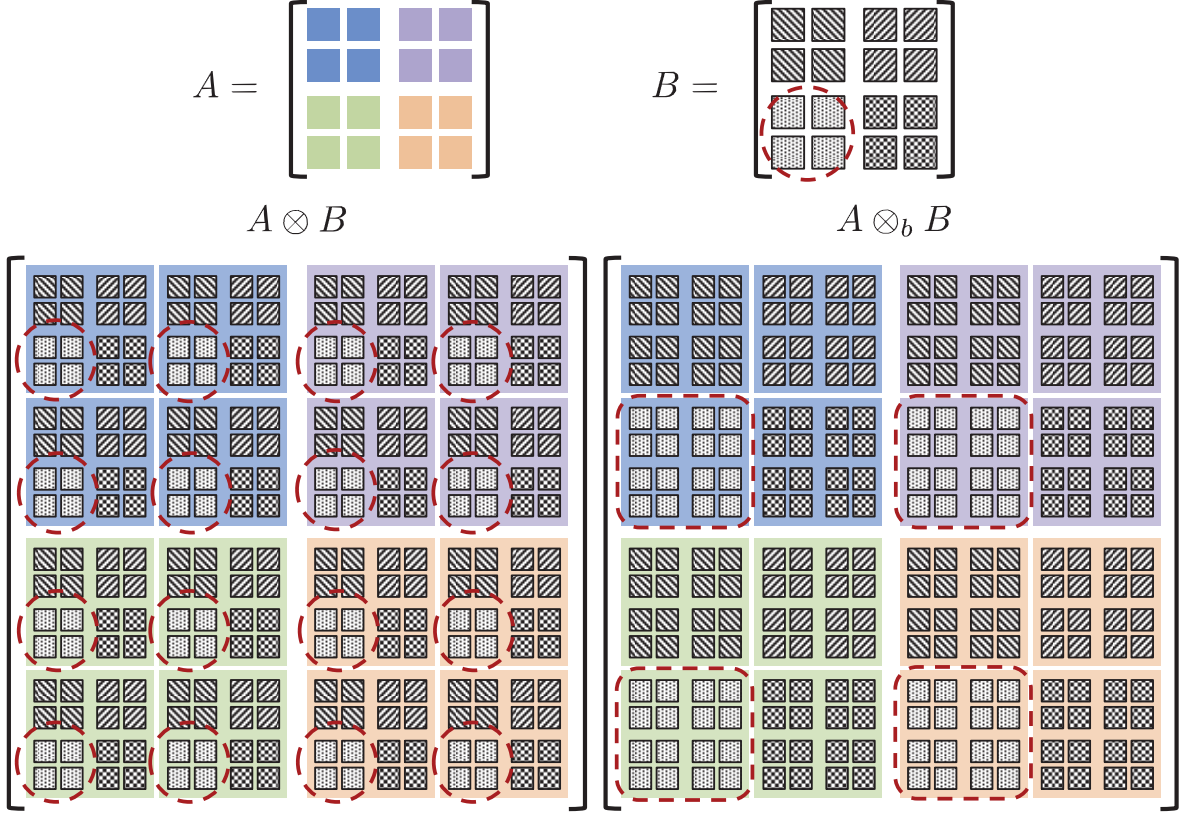


Fig. 3. Comparing two Kronecker product operations: \otimes versus \otimes_b . The operation \otimes destroys the locality of the blocks from matrix B while the operation \otimes_b preserves the locality of the blocks from both matrices A and B .

III. STEADY-STATE PERFORMANCE

When $i \rightarrow \infty$, and by the fact that \mathcal{F} is stable, we obtain from (52) that

$$\begin{aligned} z_\infty &\triangleq \lim_{i \rightarrow \infty} z_i = (I_{4M^2N^2} - \mathcal{F}^*)^{-1} \lim_{i \rightarrow \infty} y_i \\ &= (I_{4M^2N^2} - \mathcal{F}^*)^{-1} (\bar{\mathcal{A}}^\top \otimes_b \bar{\mathcal{A}}^\top + \mathcal{C}_A^\top) (\bar{\mathcal{M}} \otimes_b \bar{\mathcal{M}} + \mathcal{C}_M) \lim_{i \rightarrow \infty} \text{bvec}(\mathbb{E} \mathcal{R}_i(\mathbf{w}_{i-1})) \end{aligned} \quad (59)$$

where we also used (47) and (48). Now note that

$$\begin{aligned} \|\mathcal{R}_i(\mathbf{1}_N \otimes w^o) - \mathbb{E} \mathcal{R}_i(\mathbf{w}_{i-1})\| &\stackrel{(a)}{\leq} \mathbb{E} \|\mathcal{R}_i(\mathbf{1}_N \otimes w^o) - \mathcal{R}_i(\mathbf{w}_{i-1})\| \\ &\stackrel{(b)}{\leq} \kappa_v \cdot \mathbb{E} \|\mathbf{1}_N \otimes w^o - \mathbf{w}_{i-1}\|^{\gamma_v} \\ &= \kappa_v \cdot \mathbb{E} [\|\tilde{\mathbf{w}}_{i-1}\|^4]^{\gamma_v/4} \\ &\stackrel{(c)}{\leq} \kappa_v \cdot [\mathbb{E} \|\tilde{\mathbf{w}}_{i-1}\|^4]^{\gamma_v/4} \end{aligned} \quad (60)$$

where step (a) is by Jensen's inequality; step (b) is by (8) in Assumption 1; and step (c) is by Jensen's inequality and the fact that $|\cdot|^{\gamma_v/4}$ is concave due to $0 < \gamma_v/4 \leq 1$. Now, from Theorem 2 in Part I [2],

we know that $\limsup_{i \rightarrow \infty} \mathbb{E} \|\tilde{\mathbf{w}}_{i-1}\|^4 \leq O(\nu^2)$ under Assumption 2. Therefore, we obtain from (60) that

$$\limsup_{i \rightarrow \infty} \|\mathcal{R}_i(\mathbb{1}_N \otimes w^o) - \mathbb{E}\mathcal{R}_i(\mathbf{w}_{i-1})\| \leq O(\nu^{\gamma_v/2}) \quad (61)$$

According to (61), we can replace $\mathbb{E}\mathcal{R}_i(\mathbf{w}_{i-1})$ in (59) by $\mathcal{R}_i(\mathbb{1}_N \otimes w^o)$ with an error in the order of $\nu^{\gamma_v/2}$. Let

$$z \triangleq (I_{4M^2N^2} - \mathcal{F}^*)^{-1}(\bar{\mathcal{A}}^\top \otimes_b \bar{\mathcal{A}}^\top + \mathcal{C}_A^\top)(\bar{\mathcal{M}} \otimes_b \bar{\mathcal{M}} + \mathcal{C}_M)\text{bvec}(\mathcal{R}) \quad (62)$$

From (198) in Part I [2], we know that the second-order moments of $\{\boldsymbol{\mu}_k(i)\}$ are in the order of ν^2 . Hence, by (84) from Part I [2], (31), and (37), it is easy to verify that

$$\|\bar{\mathcal{M}} \otimes_b \bar{\mathcal{M}} + \mathcal{C}_M\| = O(\nu^2) \quad (63)$$

Using (62), (63), and the fact that $\|(I_{4M^2N^2} - \mathcal{F}^*)^{-1}\| = O(\nu^{-1})$ from Lemma 5 further ahead, we conclude that

$$\|z\| = O(\nu) \quad (64)$$

Then, by using (9) and (61)–(64), we obtain from (59) that

$$z_\infty = z + O(\nu^{1+\gamma_v/2}), \quad \|z_\infty\| = O(\nu) \quad (65)$$

Define the steady-state average network MSD by

$$\text{MSD}^{\text{net}} \triangleq \lim_{i \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N \mathbb{E} \|\tilde{\mathbf{w}}_{k,i}\|^2 \quad (66)$$

and the steady-state individual MSD for agent k by

$$\text{MSD}_k \triangleq \lim_{i \rightarrow \infty} \mathbb{E} \|\tilde{\mathbf{w}}_{k,i}\|^2 \quad (67)$$

Theorem 3 (Steady-state MSD): It holds that

$$\text{MSD}^{\text{net}} = \frac{1}{2N} z^* \text{bvec}(I_{2MN}) + O(\nu^{1+\gamma_o}) \quad (68)$$

$$\text{MSD}_k = \frac{1}{2} z^* \text{bvec}(E_{kk} \otimes I_{2M}) + O(\nu^{1+\gamma_o}) \quad (69)$$

where z is given by (62),

$$\gamma_o \triangleq \frac{1}{2} \min\{1, \gamma_v\} \quad (70)$$

and E_{kk} is the $N \times N$ basis matrix that only has one non-zero element, which is equal to 1, at the (k, k) -th entry.

Proof: From (53) by selecting $\Sigma = I_{2MN}$, and also using (59) and (65), we get

$$\begin{aligned} \lim_{i \rightarrow \infty} \mathbb{E} \|\tilde{\mathbf{w}}'_i\|^2 &= z_\infty^* \text{bvec}(I_{2MN}) \\ &= z_\infty^* \text{bvec}(I_{2MN}) + O(\nu^{1+\gamma_v/2}) \\ &= O(\nu) \end{aligned} \quad (71)$$

Likewise, by selecting $\Sigma = E_{kk} \otimes I_{2M}$, we get

$$\begin{aligned} \lim_{i \rightarrow \infty} \mathbb{E} \|\tilde{\mathbf{w}}'_i\|_{E_{kk} \otimes I_{2M}}^2 &= z_\infty^* \text{bvec}(E_{kk} \otimes I_{2M}) \\ &= z_\infty^* \text{bvec}(E_{kk} \otimes I_{2M}) + O(\nu^{1+\gamma_v/2}) \\ &= O(\nu) \end{aligned} \quad (72)$$

Note further that

$$\mathbb{E} \|\tilde{\mathbf{w}}_i\|^2 = \mathbb{E} \|\tilde{\mathbf{w}}'_i\|^2 + \mathbb{E} \|\tilde{\mathbf{w}}_i - \tilde{\mathbf{w}}'_i\|^2 + 2\Re \mathbb{E}[\tilde{\mathbf{w}}_i'^* (\tilde{\mathbf{w}}_i - \tilde{\mathbf{w}}'_i)] \quad (73)$$

Using the Cauchy-Schwartz inequality, it can be verified that

$$|\Re \mathbb{E}[\tilde{\mathbf{w}}_i'^* (\tilde{\mathbf{w}}_i - \tilde{\mathbf{w}}'_i)]| \leq \sqrt{\mathbb{E} \|\tilde{\mathbf{w}}'_i\|^2 \cdot \mathbb{E} \|\tilde{\mathbf{w}}_i - \tilde{\mathbf{w}}'_i\|^2} \quad (74)$$

From Theorem 1, we have

$$\lim_{i \rightarrow \infty} \mathbb{E} \|\tilde{\mathbf{w}}_i - \tilde{\mathbf{w}}'_i\|^2 \leq O(\nu^2) \quad (75)$$

Substituting (71) and (75) into (73), and using (74), we get

$$\begin{aligned} \lim_{i \rightarrow \infty} \mathbb{E} \|\tilde{\mathbf{w}}_i\|^2 &= \lim_{i \rightarrow \infty} \mathbb{E} \|\tilde{\mathbf{w}}'_i\|^2 + O(\nu^2) + 2\sqrt{O(\nu) \cdot O(\nu^2)} \\ &= \lim_{i \rightarrow \infty} \mathbb{E} \|\tilde{\mathbf{w}}'_i\|^2 + O(\nu^{3/2}) \end{aligned} \quad (76)$$

Results (68) and (69) follow from (71), (72), and (76). ■

Result (68) generalizes its counterpart (276) from [5] for the synchronous diffusion strategy. Since expressions (71) and (72) are both related to the vector z in (62), let us examine z more closely to reveal the implications of asynchronous adaptation and learning on performance. Theorem 4 in the following section will lead to powerful alternative expressions for (71) and (72). The new expressions will highlight some important properties about the behavior of the asynchronous network in steady-state, such as the behavior that was illustrated earlier in Fig. 1. The subsequent analysis relies on a useful low-rank factorization result.

IV. LOW-RANK FACTORIZATION

From (59) we see that the structure of z depends on the structure of the matrix $(I_{4M^2N^2} - \mathcal{F}^*)^{-1}$. In the following, we show that by retaining the dominant eigen-space of $(I_{4M^2N^2} - \mathcal{F}^*)^{-1}$, we can obtain a more revealing MSD expression than (68) that is still accurate to the order of $O(\nu^{1+\gamma_o})$.

A. Perron Eigenvectors

To proceed, we introduce the following condition on the matrix $\bar{A} \otimes \bar{A} + C_A$.

Assumption 3 (Primitiveness of $\bar{A} \otimes \bar{A} + C_A$): The matrix $\bar{A} \otimes \bar{A} + C_A$ is assumed to be primitive [6, p. 45], namely, that there exists a finite positive integer j such that all entries of $(\bar{A} \otimes \bar{A} + C_A)^j$ are positive. ■

Lemma 2 (Primitiveness of \bar{A}): The matrix \bar{A} is primitive if $\bar{A} \otimes \bar{A} + C_A$ is primitive.

Proof: See Appendix F. ■

Assumption 3 is guaranteed if the directed graph (digraph) associated with the matrix $\bar{A} \otimes \bar{A} + C_A$ is strongly-connected with at least one self-loop [6, pp. 30,34]. The digraph associated with $\bar{A} \otimes \bar{A} + C_A$ is the union of all possible digraphs associated with the realizations of $\mathbf{A}_i \otimes \mathbf{A}_i$ [7, p. 29]. Each possible digraph associated with $\mathbf{A}_i \otimes \mathbf{A}_i$ is a Kronecker graph of order 2 generated by the initiator \mathbf{A}_i [8]. Therefore, Assumption 3 amounts to an assumption that the *union* of all possible digraphs associated with the realizations of $\mathbf{A}_i \otimes \mathbf{A}_i$ is strongly-connected with at least one self-loop. As illustrated in Fig. 4, this condition still allows the digraphs associated with \mathbf{A}_i to be weakly-connected with or without self-loops or even to be disconnected. Important cases such as random gossip [9]–[13] or probabilistic diffusion [14], [15] are therefore not ruled out by this condition. It can be verified that the converse of Lemma 2 is generally not true: when the digraph associated with \bar{A} is primitive, the digraph associated with $\bar{A} \otimes \bar{A} + C_A$ does not even need to be connected.

By Lemma 3 from Part I [2] and the above Assumption 3, the matrix $\bar{A} \otimes \bar{A} + C_A$ is left-stochastic and primitive. It follows from the Perron-Frobenius theorem [6] [16] that this matrix has a unique eigenvalue at one and a pair of eigenvectors $\{\mathbb{1}_{N^2}, p\}$ with *positive* entries satisfying:

$$(\bar{A} \otimes \bar{A} + C_A) \cdot p = p, \quad p^\top \cdot \mathbb{1}_{N^2} = 1 \quad (77)$$

Likewise, the matrix \bar{A} is also left-stochastic and primitive. It has a unique eigenvalue at one and a pair of eigenvectors $\{\mathbb{1}_N, \bar{p}\}$ with positive entries satisfying:

$$\bar{A} \cdot \bar{p} = \bar{p}, \quad \bar{p}^\top \cdot \mathbb{1}_N = 1 \quad (78)$$

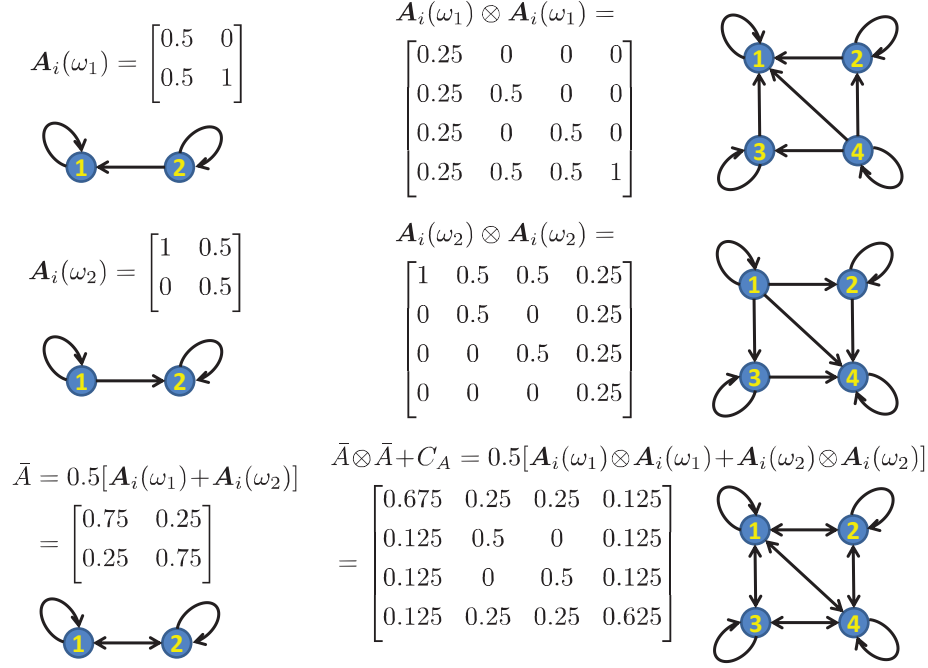


Fig. 4. An illustration of the digraph associated with $\mathbb{E}(\mathbf{A}_i \otimes \mathbf{A}_i | \mathbf{w}_{i-1}) = \bar{\mathbf{A}} \otimes \bar{\mathbf{A}} + \mathbf{C}_A$, where \mathbf{A}_i has two equally probable realizations $\{\mathbf{A}_i(\omega_1), \mathbf{A}_i(\omega_2)\}$. It can be observed that *neither* of the digraphs associated with $\mathbf{A}_i(\omega_j) \otimes \mathbf{A}_i(\omega_j)$, $j = 1, 2$, is strongly-connected due to the existence of the source and sink nodes, where information can only flow in *one* direction through the network. However, the digraph associated with $\mathbb{E}(\mathbf{A}_i \otimes \mathbf{A}_i | \mathbf{w}_{i-1})$, which is the union of the first two digraphs, is strongly-connected, where information can flow in *any* direction through the network.

All other eigenvalues of $\bar{\mathbf{A}} \otimes \bar{\mathbf{A}} + \mathbf{C}_A$ and $\bar{\mathbf{A}}$ are inside the unit circle. To simplify the presentation, we shall use the name “Perron eigenvector” to refer to the unique eigenvectors p and \bar{p} in the sequel. Since the vector p is of dimension $N^2 \times 1$, we partition it into N sub-vectors of dimension $N \times 1$ each:

$$p = \text{col}\{p_1, p_2, \dots, p_N\} \quad (79)$$

where p_k denotes the k -th sub-vector. We further define an $N \times N$ matrix P_p whose columns are the sub-vectors $\{p_k\}$:

$$P_p \triangleq \text{unvec}(p) = \begin{bmatrix} p_1 & p_2 & \dots & p_N \end{bmatrix} \quad (80)$$

We use $p_{\ell,k}$ to denote the (ℓ, k) -th element of matrix P_p , which is equal to the ℓ -th element of p_k .

Lemma 3 (Properties of P_p): The matrix P_p in (80) is symmetric positive semi-definite and it satisfies $P_p \mathbf{1}_N = \bar{p}$, where the \bar{p} is the Perron eigenvector in (78).

Proof: See Appendix G. ■

From Lemma 3, we get the following useful relations:

$$p_{\ell,k} = p_{k,\ell}, \quad \sum_{k=1}^N p_{\ell,k} = \bar{p}_\ell, \quad \sum_{\ell=1}^N p_{\ell,k} = \bar{p}_k \quad (81)$$

B. Low-Rank Approximation

We return to our earlier objective of seeking a low-rank factorization for the matrix $(I_{4M^2N^2} - \mathcal{F}^*)^{-1}$. For this purpose, we first introduce the $4M^2 \times 4M^2$ Hermitian matrix:

$$F \triangleq \sum_{k=1}^N \sum_{\ell=1}^N p_{\ell,k} [\bar{D}_\ell^\top \otimes \bar{D}_k + c_{\mu,\ell,k} (H_\ell^\top \otimes H_k)] \quad (82)$$

where \bar{D}_k is given by (32).

Lemma 4 (Spectral radius of F): The matrix F in (82) is stable if condition (23) is satisfied. Moreover,

$$\rho(F) = [1 - \lambda_{\min}(H)]^2 + O(\nu^2) = 1 - O(\nu) \quad (83)$$

where

$$H \triangleq \sum_{k=1}^N \bar{p}_k \bar{\mu}_k H_k \quad (84)$$

It can be verified that $\|H\| = O(\nu)$ and the $O(\nu^2)$ term in (83) is negligible by Assumption 2.

Proof: See Appendix H. ■

Lemma 5 (Low-rank approximation): Under Assumptions 2 and 3, it holds that

$$(I_{4N^2M^2} - \mathcal{F})^{-1} = (p\mathbf{1}_{N^2}^\top) \otimes (I_{4M^2} - F)^{-1} + O(1) \quad (85)$$

$$(p\mathbf{1}_{N^2}^\top) \otimes (I_{4M^2} - F)^{-1} = O(\nu^{-1}) \quad (86)$$

Under Assumption 2 where $\nu \ll 1$, the term in (86) dominates the $O(1)$ term in (85). Moreover,

$$\rho(\mathcal{F}) = \rho(F) + O(\nu^{1+1/N^2}) \quad (87)$$

where the $\rho(F)$ from (83) dominates the $O(\nu)$ term in (87).

Proof: See Appendix I. ■

In expression (82) we observe that the matrix F is dependent on the first and second-order moments of the random step-sizes, i.e., $\{\bar{\mu}_k\}$ and $\{c_{\mu,\ell,k}\}$, and is also dependent on the first and second order moments of the random combination coefficient matrix, i.e., \bar{A} and C_A , through the dependence on the Perron eigenvector p . Let us introduce two $4M^2 \times 4M^2$ matrices:

$$R \triangleq \sum_{k=1}^N p_{k,k} (\bar{\mu}_k^2 + c_{\mu,k,k}) R_k \quad (88)$$

$$Z \triangleq \text{unvec} \left((I_{4M^2} - F)^{-1} \text{vec}(R) \right) \quad (89)$$

where R_k is given by (10). Using (186)–(188) and (214) in Appendix I, we can verify that

$$\|R\| = O(\nu^2), \quad \|Z\| = O(\nu) \quad (90)$$

Then, using Lemma 5, we can establish the following useful result about the structure of the steady-state network error covariance matrix.

Theorem 4 (Network error covariance matrix): In steady-state, the covariance matrix of the network error $\tilde{\mathbf{w}}'_i$ from the long-term model (24) can be approximated by

$$\lim_{i \rightarrow \infty} \mathbb{E} \tilde{\mathbf{w}}'_i \tilde{\mathbf{w}}'^{*}_i = (\mathbf{1}_N \mathbf{1}_N^\top) \otimes Z + O(\nu^{1+\gamma'_o}) \quad (91)$$

where Z is from (89),

$$\gamma'_o \triangleq \frac{1}{2} \min\{2, \gamma_v\} \quad (92)$$

and the first term on the RHS is dominant.

Proof: See Appendix J. ■

According to Theorem 4, the (cross-) covariance matrices of $\{\tilde{\mathbf{w}}'_{k,i}\}$, which are uniformly expressed by $\mathbb{E} \tilde{\mathbf{w}}'_{k,i} \tilde{\mathbf{w}}'^{*}_{\ell,i} = \text{unvec}(z_i^{(\ell,k)})$ for all k and ℓ according to (58), can be approximated by Z in steady-state. However, this result is useful only if Z is a valid complex-Hessian-type matrix.

Definition 1 (Complex-Hessian-type matrices): Let X be an $M \times M$ positive semi-definite Hermitian matrix and let Y be an $M \times M$ symmetric matrix. Then, a positive semi-definite block matrix of the form

$$H \triangleq \begin{bmatrix} X & Y \\ Y^* & X^\top \end{bmatrix} \geq 0 \quad (93)$$

will be referred to as a complex-Hessian-type matrix. ■

The following result explains the reason for introducing this definition.

Lemma 6 (Complex-Hessian-type covariance matrices): Let \mathbf{x} denote an $M \times 1$ zero-mean complex random vector and let $R_x \triangleq \mathbb{E} \mathbf{x} \mathbf{x}^*$ and $R'_x \triangleq \mathbb{E} \mathbf{x} \mathbf{x}^\top$. Then, the covariance matrix of $\underline{\mathbf{x}} = \mathbb{T}(\mathbf{x})$ is given by

$$\mathbb{E} \underline{\mathbf{x}} \underline{\mathbf{x}}^* = \begin{bmatrix} \mathbb{E} \mathbf{x} \mathbf{x}^* & \mathbb{E} \mathbf{x} \mathbf{x}^\top \\ \mathbb{E} (\mathbf{x}^*)^\top \mathbf{x}^* & \mathbb{E} (\mathbf{x}^*)^\top \mathbf{x}^\top \end{bmatrix} = \begin{bmatrix} R_x & R'_x \\ R'^*_x & R_x^\top \end{bmatrix} \quad (94)$$

and this matrix is a complex-Hessian-type matrix.

Proof: It follows from comparing (94) to (93). ■

By Lemma 6, for *any* zero-mean complex random vector \mathbf{x} , the covariance matrix of $\mathbf{x} = \mathbb{T}(\mathbf{x})$ must be a complex-Hessian-type matrix. Therefore, in order to approximate $\{\text{unvec}(z_i^{(\ell,k)})\}$ by Z according to (91), we establish the following useful result for the matrix Z .

Lemma 7 (Properties of Z): The matrix Z in (89) is a positive semi-definite complex-Hessian-type matrix.

Proof: See Appendix K. ■

Using (91) and Lemmas 6 and 7, we arrive at the following result for the covariance and cross-covariance matrices of the steady-state error vectors $\{\tilde{\mathbf{w}}'_{k,i}\}$ from the long-term model (24).

Corollary 1 (Covariance and cross-covariance matrices): The steady-state (cross-) covariance matrices of individual errors $\{\tilde{\mathbf{w}}'_{k,i}\}$ from the long-term model (24) can be approximated by

$$\lim_{i \rightarrow \infty} \mathbb{E} \tilde{\mathbf{w}}'_{k,i} \tilde{\mathbf{w}}'^*_{\ell,i} = Z + O(\nu^{1+\gamma'_o}) \quad (95)$$

for all k and ℓ , where Z is given by (89) and is dominant due to (90), and γ'_o is given by (92).

Proof: By Lemma 7, the Z is a complex-Hessian-type matrix. According to Lemma 6, it is a valid covariance matrix for complex random vectors obtained via the transform $\mathbb{T}(\cdot)$. The approximation (95) then follows from Theorem 4. ■

C. Steady-State MSD

Using Corollary 1, we obtain two useful results about the steady-state MSD for asynchronous diffusion solutions.

Corollary 2 (Steady-state MSD): Based on the same assumptions as Theorem 4, the steady-state MSD (either the network MSD in (66) or the individual MSD in (67)) can be approximated by

$$\text{MSD}^{\text{net}} = \frac{1}{4} \text{Tr}(H^{-1}R) + O(\nu^{1+\gamma_o}) \quad (96)$$

$$\text{MSD}_k = \frac{1}{4} \text{Tr}(H^{-1}R) + O(\nu^{1+\gamma_o}) \quad (97)$$

where $\{H, R\}$ are given by (84) and (88), respectively, γ_o is from (70), and $\text{Tr}(H^{-1}R)$ is of the order of ν .

Proof: See Appendix L. ■

Corollary 3 (Clustered solutions): The steady-state *relative* MSD between any two agents k and ℓ , i.e., $\mathbb{E} \|\mathbf{w}_{k,i} - \mathbf{w}_{\ell,i}\|^2$, is negligible compared to their steady-state *absolute* MSD with respect to w^o , i.e.,

$$\lim_{i \rightarrow \infty} \mathbb{E} \|\mathbf{w}_{k,i} - \mathbf{w}_{\ell,i}\|^2 = O(\nu^{1+\gamma'_o}) \ll \max_k \text{MSD}_k = O(\nu) \quad (98)$$

where γ'_o is given by (92).

Proof: First, from Corollary 1, we have

$$\begin{aligned}
\lim_{i \rightarrow \infty} \mathbb{E} \|\tilde{\mathbf{w}}'_{k,i} - \tilde{\mathbf{w}}'_{\ell,i}\|^2 &= \lim_{i \rightarrow \infty} \mathbb{E} [\|\tilde{\mathbf{w}}'_{k,i}\|^2 + \|\tilde{\mathbf{w}}'_{\ell,i}\|^2 - \tilde{\mathbf{w}}'^*_{k,i} \tilde{\mathbf{w}}'_{\ell,i} - \tilde{\mathbf{w}}'^*_{\ell,i} \tilde{\mathbf{w}}'_{k,i}] \\
&= \text{Tr}(Z) + \text{Tr}(Z) - \text{Tr}(Z) - \text{Tr}(Z) + O(\nu^{1+\gamma'_o}) \\
&= O(\nu^{1+\gamma'_o})
\end{aligned} \tag{99}$$

From Theorem 1 and using (99), we get

$$\begin{aligned}
\lim_{i \rightarrow \infty} \mathbb{E} \|\tilde{\mathbf{w}}_{k,i} - \tilde{\mathbf{w}}_{\ell,i}\|^2 &= \lim_{i \rightarrow \infty} \mathbb{E} \|\tilde{\mathbf{w}}_{k,i} - \tilde{\mathbf{w}}'_{k,i} + \tilde{\mathbf{w}}'_{k,i} - \tilde{\mathbf{w}}'_{\ell,i} + \tilde{\mathbf{w}}'_{\ell,i} - \tilde{\mathbf{w}}_{\ell,i}\|^2 \\
&\leq \lim_{i \rightarrow \infty} 3\mathbb{E} (\|\tilde{\mathbf{w}}_{k,i} - \tilde{\mathbf{w}}'_{k,i}\|^2 + \|\tilde{\mathbf{w}}'_{k,i} - \tilde{\mathbf{w}}'_{\ell,i}\|^2 + \|\tilde{\mathbf{w}}'_{\ell,i} - \tilde{\mathbf{w}}_{\ell,i}\|^2) \\
&\leq O(\nu^2) + O(\nu^{1+\gamma'_o}) + O(\nu^2) \\
&\leq O(\nu^{1+\gamma'_o})
\end{aligned} \tag{100}$$

Using (6) from Part I [2], (100), and Corollary 2 completes the proof. \blacksquare

We illustrated Corollaries 2 and 3 earlier in Fig. 1. We note that if the cost functions, $\{J_k(w)\}$, are deterministic with known gradient vectors, then performance results in Corollaries 2 and 3 can still be deduced from our expressions by setting the gradient noise to zero.

V. CONCLUSION

We studied in some detail the MSE performance of *asynchronous* networks with random step-sizes, links, topologies, and combination coefficients. Assuming sufficiently small step-sizes, we showed that at steady-state, the error vector for every individual agent tends to cluster within $O(\nu^{1+\gamma_o})$ from each other, which means that the MSD performance is essentially uniform across the entire network. The result in Corollary 2 shows explicitly how the MSD performance of the network is affected by the asynchronous behavior. Quantities that relate to the first and second-order moments of the distribution of the random step-sizes and combination coefficients appear in these expressions. These results can be used to guide strategies for adjusting the combination weights and the rate at which the agents update their solutions and to ensure that the performance (in terms of MSD and rate of convergence) does not degrade below desirable levels.

APPENDIX A
PROOF OF LEMMA 1

Using Lemma 1 in Part I [2], we get from (14)–(15) that

$$\begin{aligned}
\|H_k - \mathbf{H}_{k,i-1}\| &= \left\| \int_0^1 [\nabla_{\underline{w}\underline{w}^*}^2 J_k(\underline{w}^o) - \nabla_{\underline{w}\underline{w}^*}^2 J_k(\underline{w}^o - t\tilde{\underline{w}}_{k,i-1})] dt \right\| \\
&\leq \int_0^1 \|\nabla_{\underline{w}\underline{w}^*}^2 J_k(\underline{w}^o) - \nabla_{\underline{w}\underline{w}^*}^2 J_k(\underline{w}^o - t\tilde{\underline{w}}_{k,i-1})\| dt \\
&\leq \int_0^1 \tau'_k \cdot t \cdot \|\tilde{\underline{w}}_{k,i-1}\| dt = \frac{\tau'_k}{2} \cdot \|\tilde{\underline{w}}_{k,i-1}\|
\end{aligned} \tag{101}$$

Using (101), we get from (18) that

$$\begin{aligned}
\|\mathbf{d}_{k,i}\|^2 &\leq [\mu_k(i)]^2 \cdot \|H_k - \mathbf{H}_{k,i-1}\|^2 \cdot \|\tilde{\underline{w}}_{k,i-1}\|^2 \\
&\leq [\mu_k(i)]^2 \cdot \frac{\tau_k'^2}{4} \cdot \|\tilde{\underline{w}}_{k,i-1}\|^4
\end{aligned} \tag{102}$$

Taking the expectation of both sides of (102) yields

$$\mathbb{E}\|\mathbf{d}_{k,i}\|^2 \leq \bar{\mu}_k^{(2)} \frac{\tau_k'^2}{4} \cdot \mathbb{E}\|\tilde{\underline{w}}_{k,i-1}\|^4 \tag{103}$$

From Theorem 2 of Part I [2], it holds for large enough i that

$$\mathbb{E}\|\tilde{\underline{w}}_{k,i}\|^4 \leq 2b_4^2 \cdot \nu^2 \tag{104}$$

Using the fact from (198) of Part I [2] that $\bar{\mu}_k^{(2)} \leq \nu^2$ for any k , and letting

$$\tau' \triangleq \max_k \tau'_k \tag{105}$$

we obtain from (103) and (104) that

$$\mathbb{E}\|\mathbf{d}_{k,i}\|^2 \leq 2\tau'^2 b_4^2 \cdot \nu^4 = O(\nu^4) \tag{106}$$

where a factor of 4 appeared due to the conversion $\mathbb{T}(\cdot)$ from (4) of Part I [2]. Then,

$$\begin{aligned}
\mathbb{E}\|\mathcal{A}_i^\top \mathbf{d}_i\|^2 &= \sum_{k=1}^N \mathbb{E} \left\| \sum_{\ell \in \mathcal{N}_{k,i}} \mathbf{a}_{\ell k}(i) \mathbf{d}_{\ell,i} \right\|^2 \\
&\stackrel{(a)}{\leq} \sum_{k=1}^N \mathbb{E} \left[\sum_{\ell \in \mathcal{N}_{k,i}} \|\mathbf{a}_{\ell k}(i)\| \|\mathbf{d}_{\ell,i}\|^2 \right] \\
&= \sum_{k=1}^N \sum_{\ell \in \mathcal{N}_k} \bar{a}_{\ell k} \mathbb{E}\|\mathbf{d}_{\ell,i}\|^2 \\
&\leq N \cdot \max_{\ell} \mathbb{E}\|\mathbf{d}_{\ell,i}\|^2
\end{aligned}$$

$$\stackrel{(b)}{\leq} 2N\tau'^2 b_4^2 \cdot \nu^4 \quad (107)$$

where step (a) is by using Jensen's inequality; and step (b) is by using (106).

APPENDIX B

PROOF OF THEOREM 1

We rewrite the original error recursion (16) and the long-term model (24) respectively as follows:

$$\tilde{\psi}_{k,i} = [I_{2M} - \mu_k(i)H_k]\tilde{\mathbf{w}}_{k,i-1} + \mu_k(i)\mathbf{v}_{k,i}(\mathbf{w}_{k,i-1}) + \mathbf{d}_{k,i} \quad (108)$$

$$\tilde{\mathbf{w}}_{k,i} = \sum_{\ell \in \mathcal{N}_{k,i}} \mathbf{a}_{\ell k}(i) \tilde{\psi}_{\ell,i} \quad (109)$$

and

$$\tilde{\psi}'_{k,i} = [I_{2M} - \mu_k(i)H_k]\tilde{\mathbf{w}}'_{k,i-1} + \mu_k(i)\mathbf{v}_{k,i}(\mathbf{w}_{k,i-1}) \quad (110)$$

$$\tilde{\mathbf{w}}'_{k,i} = \sum_{\ell \in \mathcal{N}_{k,i}} \mathbf{a}_{\ell k}(i) \tilde{\psi}'_{\ell,i} \quad (111)$$

with the prime notation for quantities associated with the long-term model (24). From (109) and (111), and using Jensen's inequality, the squared 2-norm of the difference between the two models is given by

$$\|\tilde{\mathbf{w}}_{k,i} - \tilde{\mathbf{w}}'_{k,i}\|^2 \leq \sum_{\ell \in \mathcal{N}_{k,i}} \mathbf{a}_{\ell k}(i) \|\tilde{\psi}_{\ell,i} - \tilde{\psi}'_{\ell,i}\|^2 \quad (112)$$

Taking the expectation of both sides yields

$$\begin{aligned} \mathbb{E}\|\tilde{\mathbf{w}}_{k,i} - \tilde{\mathbf{w}}'_{k,i}\|^2 &\leq \sum_{\ell \in \mathcal{N}_k} \bar{a}_{\ell k} \mathbb{E}\|\tilde{\psi}_{\ell,i} - \tilde{\psi}'_{\ell,i}\|^2 \\ &\leq \max_{\ell} \mathbb{E}\|\tilde{\psi}_{\ell,i} - \tilde{\psi}'_{\ell,i}\|^2 \end{aligned} \quad (113)$$

for all k . Then,

$$\max_k \mathbb{E}\|\tilde{\mathbf{w}}_{k,i} - \tilde{\mathbf{w}}'_{k,i}\|^2 \leq \max_k \mathbb{E}\|\tilde{\psi}_{k,i} - \tilde{\psi}'_{k,i}\|^2 \quad (114)$$

From (108) and (110), we have

$$\tilde{\psi}_{k,i} - \tilde{\psi}'_{k,i} = [I_{2M} - \mu_k(i)H_k](\tilde{\mathbf{w}}_{k,i-1} - \tilde{\mathbf{w}}'_{k,i-1}) + \mathbf{d}_{k,i} \quad (115)$$

Taking the expected squared 2-norm of both sides, we have

$$\begin{aligned} \mathbb{E}\|\tilde{\psi}_{k,i} - \tilde{\psi}'_{k,i}\|^2 &\leq \mathbb{E}\|[I_{2M} - \mu_k(i)H_k](\tilde{\mathbf{w}}_{k,i-1} - \tilde{\mathbf{w}}'_{k,i-1}) + \mathbf{d}_{k,i}\|^2 \\ &= \mathbb{E}\left\| \left(1-t\right) \frac{I_{2M} - \mu_k(i)H_k}{1-t} (\tilde{\mathbf{w}}_{k,i-1} - \tilde{\mathbf{w}}'_{k,i-1}) + t \frac{\mathbf{d}_{k,i}}{t} \right\|^2 \end{aligned}$$

$$\leq (1-t)^{-1} \mathbb{E} \|I_{2M} - \boldsymbol{\mu}_k(i) H_k\|^2 \cdot \mathbb{E} \|\tilde{\mathbf{w}}_{k,i-1} - \tilde{\mathbf{w}}'_{k,i-1}\|^2 + t^{-1} \mathbb{E} \|\mathbf{d}_{k,i}\|^2 \quad (116)$$

for any $0 < t < 1$, where we used Jensen's inequality in the second inequality. By condition (19), it can be verified that

$$\begin{aligned} \mathbb{E} \|I_{2M} - \boldsymbol{\mu}_k(i) H_k\|^2 &\leq 1 - 2\bar{\mu}_k^{(1)} \lambda_{k,\min} + \bar{\mu}_k^{(2)} \lambda_{k,\max}^2 \\ &\stackrel{(a)}{\leq} 1 - \bar{\mu}_k^{(1)} \lambda_{k,\min} \\ &\leq \left(1 - \frac{1}{2} \bar{\mu}_k^{(1)} \lambda_{k,\min}\right)^2 < 1 \end{aligned} \quad (117)$$

where step (a) is from (192) of Part I [2]. Substituting $t = \frac{1}{2} \bar{\mu}_k^{(1)} \lambda_{k,\min} < 1$ and (117) into (116) yields

$$\mathbb{E} \|\tilde{\boldsymbol{\psi}}_{k,i} - \tilde{\boldsymbol{\psi}}'_{k,i}\|^2 \leq \left(1 - \frac{1}{2} \bar{\mu}_k^{(1)} \lambda_{k,\min}\right) \mathbb{E} \|\tilde{\mathbf{w}}_{k,i-1} - \tilde{\mathbf{w}}'_{k,i-1}\|^2 + \frac{2}{\lambda_{k,\min} \bar{\mu}_k^{(1)}} \mathbb{E} \|\mathbf{d}_{k,i}\|^2 \quad (118)$$

Using (103), the second term on the RHS of (118) can be bounded for large enough i by

$$\begin{aligned} \frac{2}{\lambda_{k,\min} \bar{\mu}_k^{(1)}} \cdot \mathbb{E} \|\mathbf{d}_{k,i}\|^2 &\leq \frac{2}{\lambda_{k,\min} \bar{\mu}_k^{(1)}} \cdot \bar{\mu}_k^{(2)} \frac{\tau_k'^2}{4} \cdot \mathbb{E} \|\tilde{\mathbf{w}}_{k,i-1}\|^4 \\ &= \frac{\tau_k'^2}{2\lambda_{k,\min}} \cdot \frac{\bar{\mu}_k^{(2)}}{\bar{\mu}_k^{(1)}} \cdot \mathbb{E} \|\tilde{\mathbf{w}}_{k,i-1}\|^4 \\ &\leq \frac{\tau_k'^2 \nu}{2\lambda_{k,\min}} \cdot \mathbb{E} \|\tilde{\mathbf{w}}_{k,i-1}\|^4 \end{aligned} \quad (119)$$

where we used the fact from (200) of Part I [2] that $\nu \geq \bar{\mu}_k^{(2)} / \bar{\mu}_k^{(1)}$. Substituting (119) into (118) yields

$$\mathbb{E} \|\tilde{\boldsymbol{\psi}}_{k,i} - \tilde{\boldsymbol{\psi}}'_{k,i}\|^2 \leq \left(1 - \frac{1}{2} \bar{\mu}_k^{(1)} \lambda_{k,\min}\right) \mathbb{E} \|\tilde{\mathbf{w}}_{k,i-1} - \tilde{\mathbf{w}}'_{k,i-1}\|^2 + \frac{\tau_k'^2 \nu}{2\lambda_{k,\min}} \mathbb{E} \|\tilde{\mathbf{w}}_{k,i-1}\|^4 \quad (120)$$

Therefore,

$$\begin{aligned} \max_k \mathbb{E} \|\tilde{\boldsymbol{\psi}}_{k,i} - \tilde{\boldsymbol{\psi}}'_{k,i}\|^2 &\leq \max_k \left(1 - \frac{1}{2} \bar{\mu}_k^{(1)} \lambda_{k,\min}\right) \max_k \mathbb{E} \|\tilde{\mathbf{w}}_{k,i-1} - \tilde{\mathbf{w}}'_{k,i-1}\|^2 \\ &\quad + \max_k \left[\frac{\tau_k'^2 \nu}{2\lambda_{k,\min}} \cdot \mathbb{E} \|\tilde{\mathbf{w}}_{k,i-1}\|^4 \right] \end{aligned} \quad (121)$$

Substituting (121) into (114) yields

$$\max_k \mathbb{E} \|\tilde{\mathbf{w}}_{k,i} - \tilde{\mathbf{w}}'_{k,i}\|^2 \leq \gamma \cdot \max_k \mathbb{E} \|\tilde{\mathbf{w}}_{k,i-1} - \tilde{\mathbf{w}}'_{k,i-1}\|^2 + \frac{\tau'^2 \nu}{2 \min_k \lambda_{k,\min}} \cdot \max_k \mathbb{E} \|\tilde{\mathbf{w}}_{k,i-1}\|^4 \quad (122)$$

where

$$\gamma \triangleq \max_k \left(1 - \frac{1}{2} \bar{\mu}_k^{(1)} \lambda_{k,\min}\right) = 1 - \frac{1}{2} \min_k \{\bar{\mu}_k^{(1)} \cdot \lambda_{k,\min}\} \quad (123)$$

When condition (19) holds, it can be verified by using (191) from Part I [2] that $|\gamma| < 1$. Then, we get from (122) that

$$\limsup_{i \rightarrow \infty} \left[\max_k \mathbb{E} \|\tilde{\mathbf{w}}_{k,i} - \tilde{\mathbf{w}}'_{k,i}\|^2 \right] \leq \frac{\tau'^2 \nu}{(1-\gamma) \cdot 2 \min_k \lambda_{k,\min}} \cdot \limsup_{i \rightarrow \infty} \left[\max_k \mathbb{E} \|\tilde{\mathbf{w}}_{k,i-1}\|^4 \right]$$

$$\leq \frac{4\tau'^2 b_4^2 \nu^3}{\min_k \bar{\mu}_k^{(1)} \cdot \min_k \lambda_{k,\min}^2} \leq O(\nu^2) \quad (124)$$

where we used Theorem 2 from Part I [2] and the fact from (198) of Part I [2] that $\bar{\mu}_k^{(1)} = O(\nu)$.

APPENDIX C

BLOCK OPERATIONS

Consider a block matrix \mathcal{X} of size $NM \times NM$ and partition it into $N \times N$ blocks where $X_{k\ell}$ denotes its (k, ℓ) -th sub-matrix of size $M \times M$. The block vectorization of \mathcal{X} with block size $M \times M$ is defined as follows [17]:

$$\text{bvec}(\mathcal{X}) \triangleq \text{col}\{\text{vec}(X_{11}), \text{vec}(X_{21}), \dots, \text{vec}(X_{N1}), \dots, \text{vec}(X_{1N}), \text{vec}(X_{2N}), \dots, \text{vec}(X_{NN})\} \quad (125)$$

Let \mathcal{Y} denote another block matrix of size $NM \times NM$ and let $Y_{k\ell}$ denote its (k, ℓ) -th sub-matrix of size $M \times M$. Then, the block Kronecker product of \mathcal{X} and \mathcal{Y} with block size $M \times M$ is defined by [17]:

$$\mathcal{X} \otimes_b \mathcal{Y} \triangleq \begin{bmatrix} Z_{11} & Z_{12} & \dots & Z_{1N} \\ Z_{21} & Z_{22} & \dots & Z_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{N1} & Z_{N2} & \dots & Z_{NN} \end{bmatrix} \quad (126)$$

where

$$Z_{k\ell} \triangleq \begin{bmatrix} X_{k\ell} \otimes Y_{11} & X_{k\ell} \otimes Y_{12} & \dots & X_{k\ell} \otimes Y_{1N} \\ X_{k\ell} \otimes Y_{21} & X_{k\ell} \otimes Y_{22} & \dots & X_{k\ell} \otimes Y_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ X_{k\ell} \otimes Y_{N1} & X_{k\ell} \otimes Y_{N2} & \dots & X_{k\ell} \otimes Y_{NN} \end{bmatrix} \quad (127)$$

For any matrices $\{X, Y, A, B\}$ of compatible dimensions and with blocks of size $M \times M$, it holds that

$$(X \otimes A) \otimes_b (Y \otimes B) = (X \otimes Y) \otimes (A \otimes B) \quad (128)$$

where \otimes denotes the traditional Kronecker product operation. Other useful properties for the \otimes_b operation can be found in [17, pp. 176-179] and are listed here for ease of reference:

$$\text{bvec}(\mathcal{ABC}) = (\mathcal{C}^\top \otimes_b \mathcal{A}) \cdot \text{bvec}(\mathcal{B}) \quad (129)$$

$$\text{bvec}(xy^\top) = y \otimes_b x \quad (130)$$

$$\text{Tr}(\mathcal{AB}) = [\text{bvec}(\mathcal{A}^\top)]^\top \cdot \text{bvec}(\mathcal{B}) = [\text{bvec}(\mathcal{A}^*)]^* \cdot \text{bvec}(\mathcal{B}) \quad (131)$$

$$(\mathcal{A}\mathcal{C}) \otimes_b (\mathcal{B}\mathcal{D}) = (\mathcal{A} \otimes_b \mathcal{B})(\mathcal{C} \otimes_b \mathcal{D}) \quad (132)$$

$$(\mathcal{A} + \mathcal{B}) \otimes_b (\mathcal{C} + \mathcal{D}) = \mathcal{A} \otimes_b \mathcal{C} + \mathcal{B} \otimes_b \mathcal{C} + \mathcal{A} \otimes_b \mathcal{D} + \mathcal{B} \otimes_b \mathcal{D} \quad (133)$$

$$(\mathcal{A} \otimes_b \mathcal{B})^* = \mathcal{A}^* \otimes_b \mathcal{B}^* \quad (134)$$

$$(\mathcal{A} \otimes_b \mathcal{B})^\top = \mathcal{A}^\top \otimes_b \mathcal{B}^\top \quad (135)$$

for any block matrices $\{\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}\}$ and any block vectors $\{x, y\}$ with appropriate sizes.

APPENDIX D

PROOF OF THEOREM 2

From the long-term model (41), we obtain that

$$\mathbb{E}(\tilde{\mathbf{w}}'_i \tilde{\mathbf{w}}'^{*}_i | \mathbb{F}_{i-1}) = \mathbb{E}(\mathcal{B}_i \tilde{\mathbf{w}}'_{i-1} \tilde{\mathbf{w}}'^{*}_{i-1} \mathcal{B}_i^* | \mathbb{F}_{i-1}) + \mathbb{E}(\mathbf{s}_i \mathbf{s}_i^* | \mathbb{F}_{i-1}) \quad (136)$$

where the cross terms that involve \mathbf{s}_i disappear because $\mathbb{E}(\mathcal{B}_i \tilde{\mathbf{w}}'_{i-1} \mathbf{s}_i^* | \mathbb{F}_{i-1}) = 0$ by (29) and (35). Performing the block vectorization of block size $2M$ for both sides of (136), and using (129) and (130) yield

$$\mathbb{E}[(\tilde{\mathbf{w}}'^{*}_i)^\top \otimes_b \tilde{\mathbf{w}}'_i | \mathbb{F}_{i-1}] = \mathbb{E}[(\mathcal{B}_i^*)^\top \otimes_b \mathcal{B}_i][(\tilde{\mathbf{w}}'^{*}_{i-1})^\top \otimes_b \tilde{\mathbf{w}}'_{i-1}] + \mathbb{E}[(\mathbf{s}_i^*)^\top \otimes_b \mathbf{s}_i | \mathbb{F}_{i-1}] \quad (137)$$

Using (29), (132), and (135), the second term on the RHS of (137) can be expressed as

$$\mathbb{E}[(\mathbf{s}_i^*)^\top \otimes_b \mathbf{s}_i | \mathbb{F}_{i-1}] = (\bar{\mathcal{A}} \otimes_b \bar{\mathcal{A}} + \mathcal{C}_A)^\top (\bar{\mathcal{M}} \otimes_b \bar{\mathcal{M}} + \mathcal{C}_M) r_i(\mathbf{w}_{i-1}) \quad (138)$$

Substituting (51) and (138) into (137), taking the expectation with respect to \mathbb{F}_{i-1} , and then using (48), we arrive at the desired recursion (52), namely,

$$\mathbb{E}[(\tilde{\mathbf{w}}'^{*}_i)^\top \otimes_b \tilde{\mathbf{w}}'_i] = \mathcal{F}^* \mathbb{E}[(\tilde{\mathbf{w}}'^{*}_{i-1})^\top \otimes_b \tilde{\mathbf{w}}'_{i-1}] + y_i \quad (139)$$

From (51), we know that \mathcal{G} in (50) is a factor of \mathcal{F} . Hence, we use the following result to examine the stability of \mathcal{F} .

Lemma 8 (Properties of \mathcal{G}): The matrix \mathcal{G} in (50) satisfies the following properties:

- 1) Block diagonal and Hermitian matrix: it holds that

$$\mathcal{G} = \text{diag}\{G_1, G_2, \dots, G_N\} \quad (140)$$

$$G_\ell = \text{diag}\{G_{\ell,1}, G_{\ell,2}, \dots, G_{\ell,N}\} \quad (141)$$

where G_ℓ denotes the ℓ -th block on the diagonal of \mathcal{G} with block size $4M^2N \times 4M^2N$ and $G_{\ell,k}$ denotes the k -th block on the diagonal of G_ℓ with block size $4M^2 \times 4M^2$. The block $G_{\ell,k}$ is Hermitian and is given by

$$G_{\ell,k} \triangleq \bar{D}_\ell^\top \otimes \bar{D}_k + c_{\mu,\ell,k}(H_\ell^\top \otimes H_k) \quad (142)$$

where \bar{D}_k is given by (32).

2) Norms and spectral radius: it can be verified that

$$\rho(\mathcal{G}) = \max_{k,m} \{(1 - \bar{\mu}_k \lambda_{k,m})^2 + c_{\mu,k,k} \lambda_{k,m}^2\} \quad (143)$$

where $\lambda_{k,m}$ denotes the m -th eigenvalue of H_k , $m = 1, 2, \dots, 2M$.

3) Stability: if condition (23) holds, then

$$\rho(\mathcal{G}) < 1 \quad (144)$$

Proof: See Appendix E. ■

Using the fact that \mathcal{G} is block diagonal and Hermitian, and that $\bar{A} \otimes \bar{A} + C_A$ is block left-stochastic, result (153) from [4, App. A] implies that

$$\rho(\mathcal{F}) \leq \rho(\mathcal{G}) \quad (145)$$

By (144) and (145), we conclude that \mathcal{F} is stable if condition (23) holds.

APPENDIX E

PROOF OF LEMMA 8

The first property relating to the block diagonal and Hermitian structure of (140)–(142) is established by using the definition of \otimes_b and (50). Because the matrix \mathcal{D}_i is block diagonal with block size $2M \times 2M$, the block Kronecker product:

$$\mathcal{G}_i \triangleq \mathcal{D}_i^\top \otimes_b \mathcal{D}_i \quad (146)$$

is block diagonal with block size $4M^2N \times 4M^2N$ and each block is itself block diagonal with block size $4M^2 \times 4M^2$. Let us denote the ℓ -th block on the diagonal of \mathcal{G}_i with block size $4M^2N \times 4M^2N$ by

$$G_{\ell,i} \triangleq D_{\ell,i}^\top \otimes D_i \quad (147)$$

and the k -th block on the diagonal of $G_{\ell,i}$ with block size $4M^2 \times 4M^2$ by

$$G_{\ell,k,i} \triangleq D_{\ell,i}^\top \otimes D_{k,i} \quad (148)$$

where we used the fact that $D_{\ell,i}$ is Hermitian. Then, we have

$$\mathcal{G}_i = \text{diag}\{\mathbf{G}_{1,i}, \mathbf{G}_{2,i}, \dots, \mathbf{G}_{N,i}\} \quad (149)$$

$$\mathbf{G}_{\ell,i} = \text{diag}\{G_{\ell,1,i}, G_{\ell,2,i}, \dots, G_{\ell,N,i}\} \quad (150)$$

Using (50) and taking the expectation of both sides of (149) and (150), we get (140) and (141) by identifying:

$$G_\ell = \mathbb{E}[\mathbf{G}_{\ell,i}], \quad G_{\ell,k} = \mathbb{E}[G_{\ell,k,i}] \quad (151)$$

Equation (142) follows from (151), (148), (26), and (32). Since the matrices $\{G_{\ell,k}\}$ are all Hermitian, by (140) and (141), the matrix \mathcal{G} is also Hermitian.

The second property in (143) is established by using the block diagonal and Hermitian properties of \mathcal{G} to readily conclude that $\rho(\mathcal{G}) = \max_{\ell,k} \rho(G_{\ell,k})$. Furthermore, by (142), the eigenvalues of $G_{\ell,k}$ are given by

$$\lambda_{m,n}(G_{\ell,k}) = (1 - \bar{\mu}_\ell \lambda_{\ell,n})(1 - \bar{\mu}_k \lambda_{k,m}) + c_{\mu,\ell,k} \lambda_{\ell,n} \lambda_{k,m} \quad (152)$$

for any $\ell, k = 1, 2, \dots, N$, where $\lambda_{k,m}$ denotes the m -th eigenvalue of H_k and $m, n = 1, 2, \dots, 2M$. It is straightforward to verify that

$$\lambda_{m,n}(G_{\ell,k}) = \mathbb{E}[(1 - \boldsymbol{\mu}_\ell(i) \lambda_{\ell,n})(1 - \boldsymbol{\mu}_k(i) \lambda_{k,m})] \quad (153)$$

Using Cauchy-Schwarz inequality, we get

$$\begin{aligned} |\mathbb{E}[(1 - \boldsymbol{\mu}_\ell(i) \lambda_{\ell,n})(1 - \boldsymbol{\mu}_k(i) \lambda_{k,m})]| &\leq \sqrt{\mathbb{E}[(1 - \boldsymbol{\mu}_\ell(i) \lambda_{\ell,n})^2] \cdot \mathbb{E}[(1 - \boldsymbol{\mu}_k(i) \lambda_{k,m})^2]} \\ &\leq \max_{k,m} \{\mathbb{E}[(1 - \boldsymbol{\mu}_k(i) \lambda_{k,m})^2]\} \\ &= \max_{k,m} \{(1 - \bar{\mu}_k \lambda_{k,m})^2 + c_{\mu,k,k} \lambda_{k,m}^2\} \end{aligned} \quad (154)$$

where the first inequality becomes equality when $\ell = k$ and $n = m$. From (152)–(154) we get

$$|\lambda_{m,n}(G_{\ell,k})| \leq \max_{k,m} \{(1 - \bar{\mu}_k \lambda_{k,m})^2 + c_{\mu,k,k} \lambda_{k,m}^2\} \quad (155)$$

for any ℓ, k, m , and n . Since the above inequality applies to *all* eigenvalues of $G_{\ell,k}$, and since $G_{\ell,k}$ is Hermitian, we get

$$\rho(G_{\ell,k}) \leq \max_{k,m} \{(1 - \bar{\mu}_k \lambda_{k,m})^2 + c_{\mu,k,k} \lambda_{k,m}^2\} \quad (156)$$

Furthermore, from (152) we know that

$$\lambda_{m,m}(G_{k,k}) = (1 - \bar{\mu}_k \lambda_{k,m})^2 + c_{\mu,k,k} \lambda_{k,m}^2 \quad (157)$$

so that equality in (156) is achievable for some k and m .

For the third property in (144), we introduce the quadratic function

$$f(x) \triangleq (1 - \bar{\mu}_k x)^2 + c_{\mu,k,k} x^2 \quad (158)$$

with $x \in [\lambda_{k,\min}, \lambda_{k,\max}]$. It is easy to verify that $f(x)$ achieves its maximum value at either one of its boundaries:

$$\begin{aligned} f(x) &\leq \max\{f(\lambda_{k,\min}), f(\lambda_{k,\max})\} \\ &\leq 1 - 2\bar{\mu}_k \lambda_{k,\min} + (\bar{\mu}_k^2 + c_{\mu,k,k}) \lambda_{k,\max}^2 \end{aligned} \quad (159)$$

From Assumption 2 in Part I [2] we have $\lambda_{k,\min} \leq \lambda_{k,m} \leq \lambda_{k,\max}$ for any k and m . We then deduce from (159) that

$$f(\lambda_{k,m}) \leq 1 - 2\bar{\mu}_k \lambda_{k,\min} + (\bar{\mu}_k^2 + c_{\mu,k,k}) \lambda_{k,\max}^2 \quad (160)$$

for any k and m . Using (143), (158), and (160), we get

$$\begin{aligned} \rho(\mathcal{G}) &= \max_{k,m} \{f(\lambda_{k,m})\} \\ &\leq \max_k \{f(\lambda_{k,\min}), f(\lambda_{k,\max})\} \\ &< \max_k \{f(\lambda_{k,\min}), f(\lambda_{k,\max})\} + \alpha(\bar{\mu}_k^2 + c_{\mu,k,k}) \end{aligned} \quad (161)$$

where $\alpha > 0$ by Assumption 1. When condition (23) holds, using (144) from Part I [2], we have

$$\max_k \{1 - 2\bar{\mu}_k \lambda_{k,\min} + (\bar{\mu}_k^2 + c_{\mu,k,k})(\lambda_{k,\max}^2 + \alpha)\} < 1 \quad (162)$$

Therefore, by (161) and (162), if condition (23) holds, then $\rho(\mathcal{G}) < 1$, which completes the proof.

APPENDIX F

PROOF OF LEMMA 2

From Lemma 3 in Part I [2] we know that the matrices $\bar{A} \otimes \bar{A} + C_A$ and \bar{A} are both left-stochastic. To establish the desired result, we only need to show that the matrix $\bar{A} \otimes \bar{A}$ is primitive if $\bar{A} \otimes \bar{A} + C_A$ is primitive. This is because if $\bar{A} \otimes \bar{A}$ is primitive, then for some finite positive integer $j > 0$, the matrix $(\bar{A} \otimes \bar{A})^j$ has strictly positive entries. Since $(\bar{A} \otimes \bar{A})^j = \bar{A}^j \otimes \bar{A}^j$ and \bar{A} has nonnegative entries, \bar{A}^j must have strictly positive entries. Therefore, \bar{A} is primitive.

In order to prove that the matrix $\bar{A} \otimes \bar{A}$ is primitive if $\bar{A} \otimes \bar{A} + C_A$ is primitive, we first introduce the following concept.

Definition 2 (Comparing sparsity): For any two matrices $\{A, B\}$ with nonnegative entries and of the same size, the matrix A is called *sparser* than B , or, equivalently, B is called *denser* than A , if, and only if, $[B]_{\ell k} > 0$ whenever $[A]_{\ell k} > 0$ for any k and ℓ . ■

It is straightforward to verify the following three useful properties related to Definition 2.

Lemma 9 (Denser product): For any $M \times N$ matrices $\{A, B\}$ and any $N \times P$ matrices $\{C, D\}$ all with nonnegative entries, if B is denser than A and D is denser than C , then BD is denser than AC . ■

Lemma 10 (Denser Kronecker product): For any $M \times N$ matrices $\{A, B\}$ and any $P \times Q$ matrices $\{C, D\}$ all with nonnegative entries, if B is denser than A and D is denser than C , then $B \otimes D$ is denser than $A \otimes C$. ■

Lemma 11 (Sum is not denser): For any set of $M \times N$ matrices $\{A_i\}$ with nonnegative entries, where $i \in \mathcal{I}$ and \mathcal{I} is an index set (which can be uncountable), if there exists an $M \times N$ matrix B with nonnegative entries such that B is denser than every A_i , $i \in \mathcal{I}$, and assuming that the sum $S \triangleq \sum_{i \in \mathcal{I}} A_i$ exists, then B is also denser than S . ■

Now, from Lemma 2 in Part I [2], we know that \bar{A} is denser than any realization of A_i , say, $A_i(\omega)$ where $\omega \in \Omega$ and Ω is the sample space of A_i . Using Lemma 10, we get that $\bar{A} \otimes \bar{A}$ is denser than any $A_i(\omega) \otimes A_i(\omega)$. Using Lemma 11 and the fact that the probability measures only take nonnegative values, we get that $\bar{A} \otimes \bar{A}$ is denser than $\bar{A} \otimes \bar{A} + C_A = \mathbb{E}[A_i \otimes A_i]$. If $\bar{A} \otimes \bar{A} + C_A$ is primitive, then there exists a finite positive integer $j > 0$ such that $(\bar{A} \otimes \bar{A} + C_A)^j$ has strictly positive entries. Using Lemma 9, we know that $(\bar{A} \otimes \bar{A})^j$ must be denser than $(\bar{A} \otimes \bar{A} + C_A)^j$. Therefore, $(\bar{A} \otimes \bar{A})^j$ must also have strictly positive entries, which means that $\bar{A} \otimes \bar{A}$ must be primitive.

APPENDIX G

PROOF OF LEMMA 3

We first show that $P_p = P_p^\top$, or equivalently,

$$p = \text{vec}(P_p^\top) \quad (163)$$

Lemma 12 (Vec-permutation matrix): The $N^2 \times N^2$ vec-permutation matrix Π is a matrix whose columns are formed from the basis vectors in \mathbb{R}^{N^2} and it satisfies:

$$\text{vec}(A) = \Pi \cdot \text{vec}(A^\top) \quad (164)$$

for any $N \times N$ matrix A . Then, for any $N \times N$ matrices $\{A, B\}$,

$$A \otimes B = \Pi(B \otimes A)\Pi \quad (165)$$

In addition, $\Pi = \Pi^\top = \Pi^* = \Pi^{-1}$.

Proof: See [18, Tabs. I and II] [19, Eqs. (5) and (6)]. ■

Let Π be the permutation matrix that satisfies

$$\text{vec}(P_p^\top) = \Pi \cdot \text{vec}(P_p) \quad (166)$$

From (80) and (166), proving (163) is equivalent to proving

$$p = \Pi \cdot p \quad (167)$$

To establish (167), we only need to show that $\Pi \cdot p$ is the Perron eigenvector of $\bar{A} \otimes \bar{A} + C_A$. In that case, we can obtain (167) directly from the uniqueness of the Perron eigenvector, which is p . Thus, note that

$$\begin{aligned} \Pi(\bar{A} \otimes \bar{A} + C_A)\Pi &= \Pi[\mathbb{E}(\mathbf{A}_i \otimes \mathbf{A}_i)]\Pi \\ &\stackrel{(a)}{=} \mathbb{E}(\mathbf{A}_i \otimes \mathbf{A}_i) \\ &= \bar{A} \otimes \bar{A} + C_A \end{aligned} \quad (168)$$

where step (a) is by (165). Then, we deduce from (77) that

$$\Pi \cdot p = \Pi(\bar{A} \otimes \bar{A} + C_A)p = (\bar{A} \otimes \bar{A} + C_A)(\Pi \cdot p) \quad (169)$$

$$\mathbf{1}_{N^2}^\top \cdot \Pi \cdot p = \mathbf{1}_{N^2}^\top \cdot p = 1 \quad (170)$$

where we used the fact that $\Pi^2 = I_{N^2}$ by Lemma 12 and $\Pi \cdot \mathbf{1}_{N^2} = \mathbf{1}_{N^2}$. Results (169) and (170) establish that $\Pi \cdot p$ is the Perron eigenvector of $\bar{A} \otimes \bar{A} + C_A$ and proves (167).

We next establish that P_p is positive semi-definite. Note that for any vector $x \in \mathbb{R}^N$:

$$x^\top P_p x = \text{vec}(x^\top P_p x) = \frac{1}{N^2} (x^\top \otimes x^\top) p \cdot \mathbf{1}_{N^2}^\top \mathbf{1}_{N^2} \quad (171)$$

by using (80) and the fact that $\mathbf{1}_{N^2}^\top \mathbf{1}_{N^2} = N^2$. Since $\bar{A} \otimes \bar{A} + C_A = \mathbb{E}(\mathbf{A}_j \otimes \mathbf{A}_j)$, we can introduce a series of fictitious random combination matrices $\{\mathbf{A}'_j; j \geq 1\}$ such that they are mutually-independent and satisfy $\mathbb{E}(\mathbf{A}'_j \otimes \mathbf{A}'_j) = \bar{A} \otimes \bar{A} + C_A$ for any $j \geq 1$. Let $\Phi_i \triangleq \prod_{j=1}^i \mathbf{A}'_j$ for any $i \geq 1$. Then,

$$\begin{aligned} \lim_{i \rightarrow \infty} \mathbb{E}(\Phi_i \otimes \Phi_i) &\stackrel{(a)}{=} \lim_{i \rightarrow \infty} \prod_{j=1}^i \mathbb{E}(\mathbf{A}'_j \otimes \mathbf{A}'_j) \\ &= \lim_{i \rightarrow \infty} (\bar{A} \otimes \bar{A} + C_A)^i \\ &\stackrel{(b)}{=} p \cdot \mathbf{1}_{N^2}^\top \end{aligned} \quad (172)$$

where step (a) is by using the fact that the $\{\mathbf{A}'_j\}$ are mutually-independent, and step (b) is by using the Perron-Frobenius Theorem [6]. Substituting (172) into (171) and using the fact that $\mathbb{1}_{N^2} = \mathbb{1}_N \otimes \mathbb{1}_N$, we get

$$x^\top P_p x = \frac{1}{N^2} \lim_{i \rightarrow \infty} \mathbb{E} [(x^\top \Phi_i \mathbb{1}_N)^2] \geq 0 \quad (173)$$

which shows that P_p is positive semi-definite.

Now we show that $P_p \mathbb{1}_N = \bar{p}$. Note from (80) and (77) that

$$P_p = \mathbb{E} [\mathbf{A}_i \cdot P_p \cdot \mathbf{A}_i^\top] \quad (174)$$

by switching the order of $\text{unvec}(\cdot)$ and $\mathbb{E}(\cdot)$ and applying $\text{unvec}(\cdot)$ to the identity $\text{vec}(ABC) = (C^\top \otimes A) \cdot \text{vec}(B)$. Furthermore, we get from (174) that

$$P_p \cdot \mathbb{1}_N = \mathbb{E} (\mathbf{A}_i P_p \mathbf{A}_i^\top) \cdot \mathbb{1}_N = \bar{A} (P_p \cdot \mathbb{1}_N) \quad (175)$$

which implies that the vector $P_p \cdot \mathbb{1}_N$ is the Perron eigenvector of \bar{A} , which is \bar{p} . Because the Perron eigenvector is unique, by (78), equation $P_p \cdot \mathbb{1}_N = \bar{p}$ must hold.

APPENDIX H

PROOF OF LEMMA 4

We first establish that F is stable if condition (23) is satisfied. From (142) and (82), we get

$$F = \sum_{\ell=1}^N \sum_{k=1}^N p_{\ell,k} G_{\ell,k} \quad (176)$$

By (77) and (80), the elements $\{p_{\ell,k}\}$ of P_p satisfy:

$$\sum_{\ell=1}^N \sum_{k=1}^N p_{\ell,k} = 1, \quad \text{and} \quad p_{\ell,k} > 0 \quad (177)$$

Then, in terms of the 2-induced norm, we have

$$\|F\| \stackrel{(a)}{\leq} \sum_{k=1}^N \sum_{\ell=1}^N p_{\ell,k} \|G_{\ell,k}\| \stackrel{(b)}{\leq} \max_{k,\ell} \|G_{\ell,k}\| \stackrel{(c)}{=} \rho(\mathcal{G}) \quad (178)$$

where step (a) is from the triangle inequality of norms; step (b) is by using (177); and step (c) is by (143). Using (144), (178), and the fact that $\rho(F) = \|F\|$ for the Hermitian matrix F , we conclude that matrix F is stable if condition (23) holds.

We now establish expression (83). Introduce the Hermitian matrix

$$F' \triangleq \sum_{k=1}^N \sum_{\ell=1}^N \bar{p}_\ell \bar{p}_k (\bar{D}_\ell^\top \otimes \bar{D}_k) \quad (179)$$

From (32), we can rewrite F' as $F' = (I_{2M} - H)^T \otimes (I_{2M} - H)$, where we used (78) and (84). Therefore, the eigenvalues of F' are equal to the products of any two of the eigenvalues of $I_{2M} - H$, which are given by $1 - \lambda(H)$. Since $\{\bar{p}_k\}$ and $\{\bar{\mu}_k\}$ are all positive and $\{H_k\}$ are all positive definite, it is easy to verify that H in (84) is also positive definite. Then, from (84), (78), and (15), and using (8) from Part I [2] as well as Jensen's inequality [20], we get

$$0 < \lambda(H) \leq \|H\| \leq \max_k \{\bar{\mu}_k \lambda_{k,\max}\} \quad (180)$$

for all eigenvalues of H . When condition (23) holds, we get from (195) of Part I [2] that

$$\bar{\mu}_k \leq \frac{\bar{\mu}_k^{(2)}}{\bar{\mu}_k} < \frac{\lambda_{k,\min}}{\lambda_{k,\max}^2 + \alpha} < \frac{1}{\lambda_{k,\max}} \quad (181)$$

for any k . This implies that $\max_k \{\bar{\mu}_k \lambda_{k,\max}\} < 1$ and therefore, $0 < \lambda(H) < 1$ for all eigenvalues of H . From (84) and (186), we obtain

$$0 < \lambda(H) = O(\nu) < 1 \quad (182)$$

for any eigenvalue of H . Therefore, we get

$$\lambda(I_{2M} - H) = 1 - O(\nu), \quad \rho(I_{2M} - H) = 1 - \lambda_{\min}(H) \quad (183)$$

where $\lambda_{\min}(\cdot)$ denotes the smallest eigenvalue of its Hermitian matrix argument. We further get from (183) that

$$\lambda(F') = 1 - O(\nu), \quad \rho(F') = [1 - \lambda_{\min}(H)]^2 \quad (184)$$

Using Lemma 3, (77), (78), and (32), the difference between F in (82) and F' in (179) is given by

$$F - F' = \sum_{k=1}^N \sum_{\ell=1}^N \{[(p_{\ell,k} - \bar{p}_\ell \bar{p}_k) \bar{\mu}_\ell \bar{\mu}_k + p_{\ell,k} c_{\mu,\ell,k}](H_\ell^T \otimes H_k)\} \quad (185)$$

which is also Hermitian. From (198) in Part I [2], we get

$$\bar{\mu}_k \equiv \bar{\mu}_k^{(1)} \leq \nu \quad (186)$$

$$c_{\mu,k,k} \leq \bar{\mu}_k^{(2)} \leq \nu^2 \quad (187)$$

$$|c_{\mu,\ell,k}| \leq \sqrt{c_{\mu,\ell,\ell} \cdot c_{\mu,k,k}} \leq \nu^2 \quad (188)$$

where (188) is by using the Cauchy-Schwartz inequality. By (186)–(188), we get $\|F - F'\| = O(\nu^2)$.

Using a corollary of the Wielandt-Hoffman theorem [21, Corollary 8.1.6, p. 396], we then conclude that

$$|\lambda_m(F) - \lambda_m(F')| \leq \|F - F'\| = O(\nu^2) \quad (189)$$

where $\lambda_m(\cdot)$ denotes the m -th eigenvalue of its Hermitian matrix argument; the eigenvalues are assumed to be ordered from largest to smallest in each case. Result (189) implies that for every eigenvalue of F' there is an eigenvalue of F that is $O(\nu^2)$ close to it. From (189) and (184) we immediately deduce that

$$\lambda_m(F) = 1 - O(\nu), \quad \rho(F) = \rho(F') + O(\nu^2) \quad (190)$$

where $\rho(F')$ from (184) dominates the $O(\nu^2)$ term.

APPENDIX I

PROOF OF LEMMA 5

We first establish (85). Introduce the Jordan decomposition:

$$\bar{A} \otimes \bar{A} + C_A \triangleq P J Q^\top = \begin{bmatrix} p & P' \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & J' \end{bmatrix} \begin{bmatrix} \mathbb{1}_{N^2} & Q' \end{bmatrix}^\top \quad (191)$$

where J is the Jordan canonical form of $\bar{A} \otimes \bar{A} + C_A$ and J' is a sub-matrix of J containing its stable eigenvalues, P' and Q' are sub-matrices of P and Q , and $P^{-1} = Q^\top$. Then, the Jordan decomposition of $\bar{A} \otimes_b \bar{A} + C_A$ is given by

$$\bar{A} \otimes_b \bar{A} + C_A = \mathcal{P} \mathcal{J} \mathcal{Q}^\top = \begin{bmatrix} \mathcal{P}_1 & \mathcal{P}' \end{bmatrix} \begin{bmatrix} I_{4M^2} & 0 \\ 0 & \mathcal{J}' \end{bmatrix} \begin{bmatrix} \mathcal{Q}_1 & \mathcal{Q}' \end{bmatrix}^\top \quad (192)$$

where

$$\mathcal{P} \triangleq P \otimes I_{4M^2}, \quad \mathcal{P}' \triangleq P' \otimes I_{4M^2} \quad (193)$$

$$\mathcal{J} \triangleq J \otimes I_{4M^2}, \quad \mathcal{J}' \triangleq J' \otimes I_{4M^2} \quad (194)$$

$$\mathcal{Q} \triangleq Q \otimes I_{4M^2}, \quad \mathcal{Q}' \triangleq Q' \otimes I_{4M^2} \quad (195)$$

$$\mathcal{P}_1 \triangleq p \otimes I_{4M^2}, \quad \mathcal{Q}_1 \triangleq \mathbb{1}_{N^2} \otimes I_{4M^2} \quad (196)$$

Let

$$\mathcal{X} \triangleq I_{4M^2 N^2} - \mathcal{G} \quad (197)$$

where \mathcal{G} is given by (50). Then, by (51),

$$\begin{aligned} \mathcal{Q}^\top \mathcal{F} \mathcal{P} &= \mathcal{Q}^\top [\mathcal{G} \cdot (\bar{A} \otimes_b \bar{A} + C_A)] \mathcal{P} \\ &= \mathcal{Q}^\top (I_{4M^2 N^2} - \mathcal{X}) (\mathcal{P} \mathcal{J} \mathcal{Q}^\top) \mathcal{P} \\ &= \mathcal{J} - \mathcal{Q}^\top \mathcal{X} \mathcal{P} \mathcal{J} \end{aligned}$$

$$= \begin{bmatrix} I_{4M^2} - \mathcal{Q}_1^\top \mathcal{X} \mathcal{P}_1 & -\mathcal{Q}_1^\top \mathcal{X} \mathcal{P}' \mathcal{J}' \\ -\mathcal{Q}'^\top \mathcal{X} \mathcal{P}_1 & \mathcal{J}' - \mathcal{Q}'^\top \mathcal{X} \mathcal{P}' \mathcal{J}' \end{bmatrix} \quad (198)$$

From (198), we further get

$$(I_{4M^2 N^2} - \mathcal{Q}^\top \mathcal{F} \mathcal{P})^{-1} = \begin{bmatrix} \mathcal{Q}_1^\top \mathcal{X} \mathcal{P}_1 & \mathcal{Q}_1^\top \mathcal{X} \mathcal{P}' \mathcal{J}' \\ \mathcal{Q}'^\top \mathcal{X} \mathcal{P}_1 & I - \mathcal{J}' + \mathcal{Q}'^\top \mathcal{X} \mathcal{P}' \mathcal{J}' \end{bmatrix}^{-1} \quad (199)$$

where the I denotes the $4M^2(N^2 - 1) \times 4M^2(N^2 - 1)$ identity matrix. The quantity $\mathcal{Q}_1^\top \mathcal{X} \mathcal{P}_1$ in (198) can be expressed as

$$\begin{aligned} \mathcal{Q}_1^\top \mathcal{X} \mathcal{P}_1 &\stackrel{(a)}{=} \mathcal{Q}_1^\top \cdot \mathcal{P}_1 - \mathcal{Q}_1^\top \mathcal{G} \mathcal{P}_1 \\ &\stackrel{(b)}{=} (\mathbf{1}_{N^2}^\top p) \otimes I_{4M^2} - (\mathbf{1}_{N^2}^\top \otimes I_{4M^2})(\text{diag}\{G_{\ell,k}\})(p \otimes I_{4M^2}) \\ &\stackrel{(c)}{=} I_{4M^2} - F \end{aligned} \quad (200)$$

where step (a) is by (197); step (b) is by (140)–(141); and step (c) is by (176). We already know that the matrices F and \mathcal{F} are stable for sufficiently small step-sizes. Thus, the matrices $I_{4M^2 N^2} - \mathcal{F}$ and $I_{4M^2} - F$ are invertible. It follows that the quantity $\mathcal{Q}_1^\top \mathcal{X} \mathcal{P}_1$ is invertible. Moreover, the Schur complement with respect to $\mathcal{Q}_1^\top \mathcal{X} \mathcal{P}_1$ in (199) is also invertible. Let us denote the inverse of this Schur complement by

$$\Delta \triangleq [I - \mathcal{J}' + \mathcal{Q}'^\top \mathcal{X} \mathcal{P}' \mathcal{J}' - \mathcal{Q}'^\top \mathcal{X} \mathcal{P}_1 (\mathcal{Q}_1^\top \mathcal{X} \mathcal{P}_1)^{-1} \mathcal{Q}_1^\top \mathcal{X} \mathcal{P}' \mathcal{J}']^{-1} \quad (201)$$

Then, by using a formula for the inversion of block matrices [22, Eq. (7), p. 48], equality (199) can be expressed as

$$(I_{4M^2 N^2} - \mathcal{Q}^\top \mathcal{F} \mathcal{P})^{-1} = \begin{bmatrix} (\mathcal{Q}_1^\top \mathcal{X} \mathcal{P}_1)^{-1} + \Delta' & -(\mathcal{Q}_1^\top \mathcal{X} \mathcal{P}_1)^{-1} \mathcal{Q}_1^\top \mathcal{X} \mathcal{P}' \mathcal{J}' \Delta \\ -\Delta \mathcal{Q}'^\top \mathcal{X} \mathcal{P}_1 (\mathcal{Q}_1^\top \mathcal{X} \mathcal{P}_1)^{-1} & \Delta \end{bmatrix} \quad (202)$$

where

$$\Delta' \triangleq (\mathcal{Q}_1^\top \mathcal{X} \mathcal{P}_1)^{-1} \mathcal{Q}_1^\top \mathcal{X} \mathcal{P}' \mathcal{J}' \Delta \mathcal{Q}'^\top \mathcal{X} \mathcal{P}_1 (\mathcal{Q}_1^\top \mathcal{X} \mathcal{P}_1)^{-1} \quad (203)$$

Now, from (200), (82), (32), (81), and (84), we can also write

$$\mathcal{Q}_1^\top \mathcal{X} \mathcal{P}_1 = H \otimes I_{2M} + I_{2M} \otimes H - \sum_{\ell,k=1}^N p_{\ell,k} (\bar{\mu}_\ell \bar{\mu}_k + c_{\mu,\ell,k}) (H_\ell^\top \otimes H_k) \quad (204)$$

It follows from (204) and (186)–(188) that $\mathcal{Q}_1^\top \mathcal{X} \mathcal{P}_1$ is Hermitian and

$$\|\mathcal{Q}_1^\top \mathcal{X} \mathcal{P}_1\| = O(\nu), \quad \|(\mathcal{Q}_1^\top \mathcal{X} \mathcal{P}_1)^{-1}\| = O(\nu^{-1}) \quad (205)$$

Likewise, from (31) and (186)–(188), we get that

$$\|\bar{\mathcal{M}}\| = O(\nu), \quad \|\mathcal{C}_M\| = O(\nu^2) \quad (206)$$

and from (197), (50), (33), and (38), we further get

$$\|\mathcal{X}\| = \|(\bar{\mathcal{M}}\mathcal{H})^\top \otimes_b I_{2MN} + I_{2MN} \otimes_b (\bar{\mathcal{M}}\mathcal{H}) + O(\nu^2)\| = O(\nu) \quad (207)$$

since matrix \mathcal{H} is constant and independent of ν . Furthermore, it follows from (207) that

$$\|\mathcal{Q}^\top \mathcal{X} \mathcal{P}' \mathcal{J}'\| = O(\nu), \quad \|\mathcal{Q}^\top \mathcal{X} \mathcal{P}_1\| = O(\nu), \quad \|\mathcal{Q}_1^\top \mathcal{X} \mathcal{P}' \mathcal{J}'\| = O(\nu) \quad (208)$$

From (192), matrix $I - \mathcal{J}'$ is invertible and is independent of ν . Therefore,

$$\|\mathcal{J}'\| = O(1), \quad \|I - \mathcal{J}'\| = O(1), \quad \|(I - \mathcal{J}')^{-1}\| = O(1) \quad (209)$$

Then, by (201), (208), and (209), we get

$$\|\Delta\| = \|(I - \mathcal{J}' + O(\nu))^{-1}\| = O(1) \quad (210)$$

By (203), (205), (208), and (210), we further get

$$\|\Delta'\| = O(1), \quad \|(\mathcal{Q}_1^\top \mathcal{X} \mathcal{P}_1)^{-1} \mathcal{Q}_1^\top \mathcal{X} \mathcal{P}' \mathcal{J}' \Delta\| = O(1), \quad \|\Delta \mathcal{Q}^\top \mathcal{X} \mathcal{P}_1 (\mathcal{Q}_1^\top \mathcal{X} \mathcal{P}_1)^{-1}\| = O(1) \quad (211)$$

Using (211) and Assumption 2, we get from (202) that

$$(I_{4M^2N^2} - \mathcal{Q}^\top \mathcal{F} \mathcal{P})^{-1} = \begin{bmatrix} (\mathcal{Q}_1^\top \mathcal{X} \mathcal{P}_1)^{-1} & 0 \\ 0 & 0 \end{bmatrix} + O(1) \quad (212)$$

Then,

$$\begin{aligned} (I_{4M^2N^2} - \mathcal{F})^{-1} &\stackrel{(a)}{=} \mathcal{P} \begin{bmatrix} (\mathcal{Q}_1^\top \mathcal{X} \mathcal{P}_1)^{-1} & 0 \\ 0 & 0 \end{bmatrix} \mathcal{Q}^\top + O(1) \\ &\stackrel{(b)}{=} \mathcal{P}_1 (\mathcal{Q}_1^\top \mathcal{X} \mathcal{P}_1)^{-1} \mathcal{Q}_1^\top + O(1) \\ &\stackrel{(c)}{=} (p \mathbb{1}_{N^2}^\top) \otimes (I_{4M^2} - F)^{-1} + O(1) \end{aligned} \quad (213)$$

where step (a) is by using the fact that $\mathcal{P}^{-1} = \mathcal{Q}^\top$; step (b) is by using the block division in (192); step (c) is by using (196); and by (200) and (205),

$$\|(I_{4M^2} - F)^{-1}\| = O(\nu^{-1}) \quad (214)$$

Under Assumption 2, the parameter $\nu \ll 1$. Therefore, $\nu^{-1} \gg 1$ and $(p \mathbb{1}_{N^2}^\top) \otimes (I_{4M^2} - F)^{-1}$ dominates the $O(1)$ term in (213).

Finally, we establish result (87). Let

$$\mathcal{F}_s \triangleq \mathcal{Q}^\top \mathcal{F} \mathcal{P} = \begin{bmatrix} F & O(\nu) \\ O(\nu) & \mathcal{J}' + O(\nu) \end{bmatrix} \quad (215)$$

by using (198), (200), (208), and (209). Since \mathcal{F}_s is similar to \mathcal{F} , they have the same eigenvalues [22]. Since F is Hermitian, let us introduce its eigenvalue decomposition as

$$F = U \Lambda U^* \quad (216)$$

where U is a $4M^2 \times 4M^2$ unitary matrix and Λ is a $4M^2 \times 4M^2$ diagonal matrix. The $(N^2 - 1) \times (N^2 - 1)$ matrix J' , which contains the stable eigenvalues of $\bar{A} \otimes \bar{A} + C_A$ in (191), can be generally expressed as

$$J' = \begin{bmatrix} \lambda_{a,2} & & T' \\ & \ddots & \\ 0 & & \lambda_{a,N^2} \end{bmatrix} \quad (217)$$

where $\{\lambda_{a,n}\}$ are the eigenvalues of $\bar{A} \otimes \bar{A} + C_A$ with $\lambda_{a,1} = 1$ and $|\lambda_{a,n}| < 1$ for all $n = 2, 3, \dots, N^2$. In (217), the elements in the strictly upper triangular region T' are either 1 or 0, which depend on the Jordan blocks in J' . Using (217) and (194), we can express the $(2, 2)$ block in (215) as

$$\mathcal{J}' + O(\nu) = \begin{bmatrix} \lambda_{a,2} I_{4M^2} + O(\nu) & \mathcal{T}' + O(\nu) \\ & \ddots \\ O(\nu) & \lambda_{a,N^2} I_{4M^2} + O(\nu) \end{bmatrix} \quad (218)$$

where the elements in the strictly upper triangular region \mathcal{T}' are either 1 or 0, which depend on the elements of T' in (217). We now apply a similarity transformation to \mathcal{F} by multiplying

$$\mathcal{D} \triangleq \text{diag}\{\nu^\epsilon U, \nu^{2\epsilon} I_{4M^2}, \nu^{3\epsilon} I_{4M^2}, \dots, \nu^{N^2\epsilon} I_{4M^2}\} \quad (219)$$

and its inverse \mathcal{D}^{-1} on either side of (215), where $\epsilon = 1/N^2$. Using (215) and (218), we end up with

$$\mathcal{D}^{-1} \mathcal{F}_s \mathcal{D} = \left[\begin{array}{c|ccc} \Lambda & & & O(\nu^{1+\epsilon}) \\ \hline & \lambda_{a,2} I_{4M^2} + O(\nu) & & O(\nu^\epsilon) \\ & & \ddots & \\ O(\nu^\epsilon) & & & \\ & O(\nu^\epsilon) & & \lambda_{a,N^2} I_{4M^2} + O(\nu) \end{array} \right] \quad (220)$$

From (220), we know that all off-diagonal entries of $\mathcal{D}^{-1}\mathcal{F}_s\mathcal{D}$ are *at least* of the order of ν^ϵ . Therefore, using Gershgorin Theorem [21, p. 320] under Assumption 2, and since \mathcal{F} and \mathcal{F}_s have the same eigenvalues due to similarity, we get

$$|\lambda(\mathcal{F}) - \lambda(F)| \leq O(\nu^{1+\epsilon}) \text{ or } |\lambda(\mathcal{F}) - \lambda_{a,k}| \leq O(\nu^\epsilon) \quad (221)$$

where $\lambda(\mathcal{F})$ denotes the eigenvalue of \mathcal{F} and $k = 2, 3, \dots, N^2$. Result (221) implies that the eigenvalues of \mathcal{F} are either located in the Gershgorin circles that are centered at the eigenvalues of F with radii $O(\nu^{1+\epsilon})$ or in the Gershgorin circles that are centered at $\{\lambda_{a,k}; k = 2, 3, \dots, N^2\}$ with radii $O(\nu^\epsilon)$. From (190), we have

$$\rho(F) = 1 - O(\nu) < 1 \quad (222)$$

By Assumption 3 and Perron-Frobenius Theorem [6], we have

$$\rho(J') \triangleq \max_{k=2,3,\dots,N^2} |\lambda_{a,k}| < 1 \quad (223)$$

By Assumption 2, if the parameter ν is small enough to satisfy

$$O(\nu^\epsilon) + O(\nu) < 1 - \rho(J') \quad (224)$$

such that the inequality

$$\rho(J') + O(\nu^\epsilon) < 1 - O(\nu) = \rho(F) \quad (225)$$

holds, then the Gershgorin circles centered at the eigenvalues of F are isolated from those centered at $\{\lambda_{a,k}; k = 2, 3, \dots, N^2\}$. According to Gershgorin Theorem [23, p. 181], there are precisely $4M^2$ eigenvalues of \mathcal{F} satisfying

$$|\lambda(\mathcal{F}) - \lambda(F)| \leq O(\nu^{1+\epsilon}) \quad (226)$$

while all the other eigenvalues satisfy

$$|\lambda(\mathcal{F}) - \lambda_{a,k}| \leq O(\nu^\epsilon), \quad k = 2, 3, \dots, N^2 \quad (227)$$

By (225), the eigenvalues $\lambda(\mathcal{F})$ satisfying (226) are greater than those satisfying (227) in magnitude. Furthermore, when ν is sufficiently small, the Gershgorin circles centered at $\lambda_{\max}(F)$ with radius $O(\nu^{1+\epsilon})$ will become disjoint from the other circles (see Fig. 5). Then, by using (222) and Gershgorin Theorem again, we conclude from (226) that

$$\rho(\mathcal{F}) = \rho(F) + O(\nu^{1+\epsilon}) \quad (228)$$

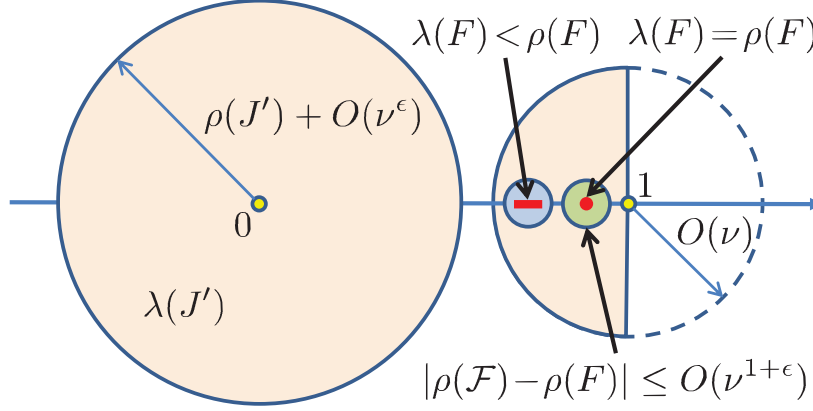


Fig. 5. An illustration of the locations of the eigenvalues of \mathcal{F} . The eigenvalues of J' are all in the left big circle, so the eigenvalues of \mathcal{F} satisfying (227) are also in the left big circle. The eigenvalues of F are all in the right big circle, so the eigenvalues of \mathcal{F} satisfying (226) are also in the right big circle. Specifically, the eigenvalues of F with $\lambda(F) < \rho(F)$ are all on the red segment on the horizontal line, so the eigenvalues of \mathcal{F} that satisfy (226) are all in the small blue circle on the left; the eigenvalues of F with $\lambda(F) = \rho(F)$ are on the red dot on the horizontal line, so the eigenvalues of \mathcal{F} that satisfy (226) are all in the small green circle on the right.

It is worth noting that from (83) we get

$$\rho(\mathcal{F}) = 1 - O(\nu) + O(\nu^{1+\epsilon}) < 1 \quad (229)$$

for $\nu \ll 1$ because $\epsilon = 1/N^2 > 1$.

APPENDIX J PROOF OF THEOREM 4

From (192) and (213), we first have

$$\begin{aligned} (\bar{\mathcal{A}} \otimes_b \bar{\mathcal{A}} + \mathcal{C}_A)(I_{4M^2N^2} - \mathcal{F})^{-1} &= \begin{bmatrix} \mathcal{P}_1 & \mathcal{P}' \end{bmatrix} \begin{bmatrix} I_{4M^2} & 0 \\ 0 & \mathcal{J}' \end{bmatrix} \begin{bmatrix} (\mathcal{Q}_1^\top \mathcal{X} \mathcal{P}_1)^{-1} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathcal{Q}_1 & \mathcal{Q}' \end{bmatrix}^\top + O(1) \\ &= \mathcal{P}_1(\mathcal{Q}_1^\top \mathcal{X} \mathcal{P}_1)^{-1} \mathcal{Q}_1^\top + O(1) \end{aligned} \quad (230)$$

Since $\mathcal{P}_1(\mathcal{Q}_1^\top \mathcal{X} \mathcal{P}_1)^{-1} \mathcal{Q}_1^\top = O(\nu^{-1})$ by (205), the first term on the RHS of (230) dominates the second term under Assumption 2. By (31), (9), (37), and (125), we get

$$\begin{aligned} \mathcal{P}_1^\top (\bar{\mathcal{M}} \otimes_b \bar{\mathcal{M}} + \mathcal{C}_M) \text{bvec}(\mathcal{R}) &= (p^\top \otimes I_{4M^2})[(\bar{\mathcal{M}} \otimes \bar{\mathcal{M}} + \mathcal{C}_M) \otimes I_{4M^2}] \text{bvec}(\mathcal{R}) \\ &= \text{vec}(R) \end{aligned} \quad (231)$$

where R is defined by (88) and is of the order of ν^2 . We then get a low-rank expression for z in (62):

$$\begin{aligned}
z &\stackrel{(a)}{=} [\mathcal{P}_1(\mathcal{Q}_1^\top \mathcal{X} \mathcal{P}_1)^{-1} \mathcal{Q}_1^\top + O(1)]^* (\bar{\mathcal{M}} \otimes_b \bar{\mathcal{M}} + \mathcal{C}_M) \text{bvec}(\mathcal{R}) \\
&\stackrel{(b)}{=} \mathcal{Q}_1 [(\mathcal{Q}_1^\top \mathcal{X} \mathcal{P}_1)^{-1}]^* \mathcal{P}_1^\top (\bar{\mathcal{M}} \otimes_b \bar{\mathcal{M}} + \mathcal{C}_M) \text{bvec}(\mathcal{R}) + O(\nu^2) \\
&\stackrel{(c)}{=} \mathcal{Q}_1 [(I_{4M^2} - F)^{-1}]^* \text{vec}(R) + O(\nu^2) \\
&\stackrel{(d)}{=} \mathbb{1}_{N^2} \otimes [(I_{4M^2} - F)^{-1} \text{vec}(R)] + O(\nu^2)
\end{aligned} \tag{232}$$

where step (a) is by (230); step (b) is by (63); step (c) is by (200) and (231); and step (d) is by (196) and the fact that F is Hermitian. The first term on the RHS of (232) is dominant due to (214) and (90). Applying $\text{unbvec}(\cdot)$ to both sides of (232) and using (89) yields

$$\text{unbvec}(z) = (\mathbb{1}_N \mathbb{1}_N^\top) \otimes Z + O(\nu^2) \tag{233}$$

where Z is given by (89) and is of the order of ν by (90). From (49) and (59), we know that $\text{unbvec}(z_\infty)$ is the steady-state covariance matrix of $\tilde{\mathbf{w}}'_i$. Using (65) and (233), the steady-state covariance matrix of $\tilde{\mathbf{w}}'_i$ can be approximated by

$$\begin{aligned}
\lim_{i \rightarrow \infty} \mathbb{E} \tilde{\mathbf{w}}'_i \tilde{\mathbf{w}}'^{\prime*}_i &= \text{unbvec}(z_\infty) \\
&= (\mathbb{1}_N \mathbb{1}_N^\top) \otimes Z + O(\nu^{1+\min\{2, \gamma_v\}/2}) \\
&= (\mathbb{1}_N \mathbb{1}_N^\top) \otimes Z + O(\nu^{1+\gamma'_o})
\end{aligned} \tag{234}$$

where γ'_o is given by (92), and the first term on the RHS is dominant due to (90).

APPENDIX K

PROOF OF LEMMA 7

From Lemma 4 and Assumption 2, we know that matrix F is stable. From (89), we get

$$Z = \text{unvec} \left(\sum_{j=0}^{\infty} F^j \text{vec}(R) \right) = \sum_{j=0}^{\infty} \text{unvec} (F^j \text{vec}(R)) \tag{235}$$

Let

$$R^{(j)} \triangleq \text{unvec} (F^j \text{vec}(R)), \quad j \geq 0 \tag{236}$$

where $R^{(0)} = R$. Then, since F is stable, the $2M \times 2M$ matrix sequence $\{R^{(j)}; j \geq 0\}$ converges to zero. Substituting (236) into (235) yields

$$Z = \sum_{j=0}^{\infty} R^{(j)} \tag{237}$$

Lemma 13 (Condition for complex-Hessian-type matrices): A sufficient and necessary condition for any $2M \times 2M$ positive semi-definite matrix H to be a complex-Hessian-type matrix in Definition 1 is to require $LH^\top L = H$, where

$$L \triangleq \begin{bmatrix} 0 & I_M \\ I_M & 0 \end{bmatrix} \quad (238)$$

satisfies $L = L^\top = L^{-1}$.

Proof: Let the $2M \times 2M$ positive semi-definite matrix be

$$H = \begin{bmatrix} A & B \\ B^* & D \end{bmatrix} \quad (239)$$

where $\{A, B, D\}$ are $M \times M$ submatrices satisfying $A = A^*$ and $D = D^*$. Then,

$$LH^\top L \triangleq \begin{bmatrix} D^\top & B^\top \\ (B^*)^\top & A^\top \end{bmatrix} \quad (240)$$

By Definition 1, the matrix H is a complex-Hessian-type matrix if, and only if, $A = D^\top$ and $B = B^\top$. It is straightforward to verify that these conditions are equivalent to the equality $LH^\top L = H$. ■

Using Lemma 13, it is easy to verify that if each $R^{(j)}$, $j \geq 0$, in (237) is Hermitian positive semi-definite and of complex-Hessian-type, then so is Z . Now, from (236), we have

$$\text{vec}(R^{(j)}) = F^j \text{vec}(R) = F F^{j-1} \text{vec}(R) = F \text{vec}(R^{(j-1)}) \quad (241)$$

From (82) and using the property $\text{vec}(ABC) = (C^\top \otimes A) \text{vec}(B)$, we get the following recursion:

$$R^{(j)} = \sum_{k=1}^N \sum_{\ell=1}^N p_{\ell,k} [\bar{D}_k R^{(j-1)} \bar{D}_\ell + c_{\mu,\ell,k} H_k R^{(j-1)} H_\ell] \quad (242)$$

We can now verify by mathematical induction that each $R^{(j)}$ is Hermitian positive semi-definite and of complex-Hessian-type. Obviously, from (88), we know that $R^{(0)} = R$ is Hermitian positive semi-definite and of complex-Hessian-type. Now, assuming that $R^{(j-1)}$ is Hermitian positive semi-definite and of complex-Hessian-type, let us verify that the same applies to $R^{(j)}$.

Since $\{\bar{D}_k, H_k\}$ are all Hermitian matrices, it is easy to verify that

$$\sum_{k=1}^N \sum_{\ell=1}^N p_{\ell,k} \bar{D}_k R^{(j-1)} \bar{D}_\ell = \begin{bmatrix} \bar{D}_1 \\ \vdots \\ \bar{D}_N \end{bmatrix}^* \begin{bmatrix} p_{1,1} R^{j-1} & \dots & p_{1,N} R^{j-1} \\ \vdots & \ddots & \vdots \\ p_{N,1} R^{j-1} & \dots & p_{N,N} R^{j-1} \end{bmatrix} \begin{bmatrix} \bar{D}_1 \\ \vdots \\ \bar{D}_N \end{bmatrix}$$

$$= (\mathbf{1}_N \otimes I_{2M})^* \bar{\mathcal{D}}^* [P_p \otimes R^{(j-1)}] \bar{\mathcal{D}} (\mathbf{1}_N \otimes I_{2M}) \quad (243)$$

and

$$\begin{aligned} \sum_{k=1}^N \sum_{\ell=1}^N p_{\ell,k} c_{\mu,\ell,k} H_k R^{(j-1)} H_\ell &= \begin{bmatrix} H_1 \\ \vdots \\ H_N \end{bmatrix}^* \begin{bmatrix} p_{1,1} c_{\mu,1,1} R^{j-1} & \cdots & p_{1,N} c_{\mu,1,N} R^{j-1} \\ \vdots & \ddots & \vdots \\ p_{N,1} c_{\mu,N,1} R^{j-1} & \cdots & p_{N,N} c_{\mu,N,N} R^{j-1} \end{bmatrix} \begin{bmatrix} H_1 \\ \vdots \\ H_N \end{bmatrix} \\ &= (\mathbf{1}_N \otimes I_{2M})^* \mathcal{H}^* [(P_p \odot C_\mu) \otimes R^{(j-1)}] \mathcal{H} (\mathbf{1}_N \otimes I_{2M}) \end{aligned} \quad (244)$$

where $\bar{\mathcal{D}}$ is from (33), \mathcal{H} is from (24), $C_\mu \triangleq [c_{\mu,\ell,k}]_{\ell,k=1}^N$, and \odot denotes the Hadamard product (the element-wise product) of matrices [24]. Since P_p is Hermitian positive semi-definite by Lemma 3, and $R^{(j-1)}$ is also Hermitian positive semi-definite by the induction hypothesis, the Kronecker product $P_p \otimes R^{(j-1)}$ must be Hermitian positive semi-definite [24, p. 245]. Therefore, the term on the LHS of (243) must be Hermitian positive semi-definite. From (43) of Part I [2], it is obvious that the C_μ is the covariance matrix of $\{\boldsymbol{\mu}_k(i)\}$ and it must be Hermitian positive semi-definite. Since P_p and C_μ are both Hermitian positive semi-definite, the Hadamard product $P_p \odot C_\mu$ is also Hermitian positive semi-definite by the Schur product Theorem [24, p. 309]. Then, the Kronecker product $(P_p \odot C_\mu) \otimes R^{(j-1)}$ must be Hermitian positive semi-definite [24, p. 245], and in turn, the term on the LHS of (244) must also be Hermitian positive semi-definite. From (243) and (244), we conclude that the $R^{(j)}$ in (242) must be Hermitian positive semi-definite.

Finally, we show that if $R^{(j-1)}$ is of complex-Hessian-type, then so is $R^{(j)}$. It is easy to verify that $\{\bar{D}_k\}$ in (32) and $\{H_k\}$ in (15) are all of complex-Hessian-type such that

$$L \bar{D}_k L = \bar{D}_k, \quad L H_k L = H_k, \quad k = 1, 2, \dots, N \quad (245)$$

From (236), we have

$$\begin{aligned} L R^{(j)} L &\stackrel{(a)}{=} \sum_{k=1}^N \sum_{\ell=1}^N p_{\ell,k} [(L \bar{D}_k L) (L R^{(j-1)} L) (L \bar{D}_\ell L) + c_{\mu,\ell,k} (L H_k L) (L R^{(j-1)} L) (L H_\ell L)] \\ &\stackrel{(b)}{=} \sum_{k=1}^N \sum_{\ell=1}^N p_{\ell,k} [\bar{D}_k R^{(j-1)} \bar{D}_\ell + c_{\mu,\ell,k} H_k R^{(j-1)} H_\ell] \\ &= R^{(j)} \end{aligned} \quad (246)$$

where step (a) is by the fact that $LL = I_{2M}$ and step (b) is by (245) and the induction hypothesis. Therefore, the matrix $R^{(j)}$ is also of complex-Hessian-type.

APPENDIX L

PROOF OF COROLLARY 2

Following an argument similar to the proof of Theorem 3, we can obtain

$$\lim_{i \rightarrow \infty} \mathbb{E} \|\tilde{\mathbf{w}}_{k,i}\|^2 = \lim_{i \rightarrow \infty} \mathbb{E} \|\tilde{\mathbf{w}}'_{k,i}\|^2 + O(\nu^{3/2}) \quad (247)$$

From (69), (247), and Corollary 1, and using the fact that $\gamma_o = \min\{1/2, \gamma'_o\}$, we can express the individual MSD by

$$\text{MSD}_k = \frac{1}{2} \text{Tr}(Z) + O(\nu^{1+\gamma_o}) \quad (248)$$

where the first term on the RHS is of the order of ν and dominates the other term. Then, by (66), we immediately get

$$\text{MSD}^{\text{net}} = \frac{1}{N} \sum_{k=1}^N \text{MSD}_k = \frac{1}{2} \text{Tr}(Z) + O(\nu^{1+\gamma_o}) \quad (249)$$

Let

$$S \triangleq H^\top \otimes I_{2M} + I_{2M} \otimes H = O(\nu) \quad (250)$$

$$Y \triangleq \sum_{\ell,k=1}^N p_{\ell,k} (\bar{\mu}_\ell \bar{\mu}_k + c_{\mu,\ell,k}) (H_\ell^\top \otimes H_k) = O(\nu^2) \quad (251)$$

by (186)–(188), where H is given by (84). It is worth noting that S is invertible when condition (19) holds and Y is always invertible by Assumption 2 in Part I [2]. By using (200), (204), (250), and (251), we get

$$I_{4M^2} - F = S + Y \quad (252)$$

Using the matrix inversion lemma [22], we get from (252) that

$$(I_{4M^2} - F)^{-1} = S^{-1} - S^{-1}(Y^{-1} + S^{-1})^{-1}S^{-1} \quad (253)$$

By (250) and (251), we know that $\|S^{-1}\| = O(\nu^{-1})$ and $\|Y^{-1}\| = O(\nu^{-2})$. Then,

$$(I_{4M^2} - F)^{-1} = S^{-1} + O(1) \quad (254)$$

By (88), we have $\|R\| = O(\nu^2)$. Using (89) and (254), we get

$$\begin{aligned} \text{Tr}(Z) &= [\text{vec}(Z)]^* \text{vec}(I_{2M}) \\ &= [\text{vec}(R)]^* (I_{4M^2} - F)^{-1} \text{vec}(I_{2M}) \\ &= [\text{vec}(R)]^* S^{-1} \text{vec}(I_{2M}) + O(\nu^2) \end{aligned} \quad (255)$$

where the first term on the RHS is of the order of ν and, therefore, is the dominant term. To further simplify (255), we consider the Lyapunov equation with respect to the unknown matrix $X \in \mathbb{C}^{2M \times 2M}$: $XH + HX = I_{2M}$, where H is given by (84). By applying the $\text{vec}(\cdot)$ operation to both sides, the Lyapunov equation is equivalent to the linear equation: $S \cdot \text{vec}(X) = \text{vec}(I_{2M})$, where $\text{vec}(X) \in \mathbb{C}^{4M^2 \times 1}$. Since S is invertible, the Lyapunov equation has a unique solution, which is given by $X = \frac{1}{2}H^{-1}$, or $\text{vec}(X) = S^{-1}\text{vec}(I_{2M}) = \frac{1}{2}\text{vec}(H^{-1})$. It then follows from (255) that

$$\text{Tr}(Z) = \frac{1}{2}\text{Tr}(H^{-1}R) + O(\nu^2) \quad (256)$$

where the first term on the RHS is of the order of ν and dominates the other term. Substituting (256) into (248) and (249) completes the proof.

REFERENCES

- [1] X. Zhao and A. H. Sayed, “Asynchronous diffusion adaptation over networks,” in *Proc. European Signal Process. Conf. (EUSIPCO)*, Bucharest, Romania, Aug. 2012, pp. 27–31.
- [2] X. Zhao and A. H. Sayed, “Asynchronous adaptation and learning over networks — Part I: Modeling and stability analysis,” *IEEE Trans. Signal Process.*, vol. xx, no. xx, pp. xxxx–xxxx, xxx 2015.
- [3] J. Chen and A. H. Sayed, “On the learning behavior of adaptive networks,” *submitted for publication*, also available as arXiv:1312.7581 [cs.MA], Dec. 2013.
- [4] X. Zhao, S.-Y. Tu, and A. H. Sayed, “Diffusion adaptation over networks under imperfect information exchange and non-stationary data,” *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3460–3475, July 2012.
- [5] A. H. Sayed, “Diffusion adaptation over networks,” in *Academic Press Library in Signal Processing*, R. Chellapa and S. Theodoridis, Eds., vol. 3, pp. 323–454. Academic Press, Elsevier, 2014.
- [6] A. Berman and R. J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*, SIAM, PA, 1994.
- [7] J. A. Bondy and U. S. R. Murty, *Graph Theory*, Springer, 2008.
- [8] J. Leskovec, D. Chakrabarti, J. Kleinberg, C. Faloutsos, and Z. Ghahramani, “Kronecker graphs: An approach to modeling networks,” vol. 11, pp. 985–1042, Sept. 2010.
- [9] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, “Randomized gossip algorithms,” *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2508–2530, June 2006.
- [10] T. C. Aysal, A. D. Sarwate, and A. G. Dimakis, “Reaching consensus in wireless networks with probabilistic broadcast,” in *Proc. Allerton Conf. Commun., Control, Comput.*, Allerton House, IL, Sept. and Oct. 2009, pp. 732–739.
- [11] T. C. Aysal, M. E. Yildiz, and A. Scaglione, “Broadcast gossip algorithms for consensus,” *IEEE Trans. Signal Process.*, vol. 57, pp. 2748–2761, 2009.
- [12] D. Jakovetic, J. Xavier, and J. M. F. Moura, “Cooperative convex optimization in networked systems: augmented Lagrangian algorithms with directed gossip communication,” *IEEE Trans. Signal Process.*, vol. 59, no. 8, pp. 3889–3902, Aug. 2011.
- [13] S. Kar and J. M. F. Moura, “Convergence rate analysis of distributed gossip (linear parameter) estimation: Fundamental limits and tradeoffs,” *IEEE J. Sel. Top. Signal Process.*, vol. 5, no. 4, pp. 674–690, Aug. 2011.
- [14] C. G. Lopes and A. H. Sayed, “Diffusion adaptive networks with changing topology,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Las Vegas, NV, Mar./Apr. 2008, pp. 3285–3288.

- [15] N. Takahashi and I. Yamada, “Link probability control for probabilistic diffusion least-mean squares over resource-constrained networks,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Dallas, TX, Mar. 2010, pp. 3518–3521.
- [16] S. U. Pillai, T. Suel, and S. Cha, “The Perron-Frobenius theorem: Some of its applications,” *IEEE Signal Process. Mag.*, vol. 22, no. 2, pp. 62–75, Mar. 2005.
- [17] R. H. Koning, H. Neudecker, and T. Wansbeek, “Block Kronecker products and the vecb operator,” *Linear Algebra Appl.*, vol. 149, pp. 165–184, Apr. 1991.
- [18] J. W. Brewer, “Kronecker products and matrix calculus in system theory,” *IEEE Trans. Circuits Syst.*, vol. 25, no. 9, pp. 772–781, Sept. 1978.
- [19] H. V. Henderson and S. R. Searle, “The vec-permutation matrix, the vec operator and Kronecker products: A review,” *Linear Multilinear Algebra*, vol. 9, no. 4, pp. 271–288, 1981.
- [20] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge Univ. Press, Cambridge, UK, 2004.
- [21] G. H. Golub and C. F. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, MD, 3rd edition, 1996.
- [22] A. J. Laub, *Matrix Analysis for Scientists and Engineers*, SIAM, PA, 2005.
- [23] G. W. Stewart and J. Sun, *Matrix Perturbation Theory*, Academic Press, Boston, MA, 1990.
- [24] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*, Cambridge Univ. Press, Cambridge, UK, 1991.